Nonparametric Estimation for Stochastic Delay Differential Equations

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)
im Fach Mathematik

eingereicht an der
Mathematisch–Naturwissenschaftlichen Fakultät II
der Humboldt–Universität zu Berlin

von

Diplom–Mathematiker

Markus Reiß

geboren am 22. Mai 1973 in Berlin

Präsident der Humboldt–Universität zu Berlin:
Prof. Dr. Jürgen Mlynek

Dekan der Mathematisch–Naturwissenschaftlichen Fakultät II:
Prof. Dr. Elmar Kulke

Gutachter:
1. Prof. Dr. Uwe Küchler, Berlin
2. Prof. Dr. Dominique Picard, Paris
3. Prof. Dr. Michael Scheutzow, Berlin

eingereicht am:	19. Oktober 2001
Tag der mündlichen Prüfung:	13. Februar 2002

Abstract

Let $(X(t),\, t \geq -r)$ be a stationary stochastic process solving the affine stochastic delay differential equation

$$dX(t) = \left(\int_{-r}^{0} X(t+s)\, da(s) \right) dt + \sigma\, dW(t), \quad t \geq 0,$$

with $r \geq 0$, $\sigma > 0$, $(W(t),\, t \geq 0)$ a standard one-dimensional Brownian motion and with a finite signed measure a. Assume that a trajectory $(X(t),\, -r \leq t \leq T)$ is observed up to time $T > 0$. In this setting the nonparametric estimation of the weight function g is considered, where $da(s) = g(s)\, ds$ is supposed to hold.

By exhibiting a close relationship with an ill-posed inverse problem, we are able to use the Galerkin projection method for the construction of an estimator of g. We regard an L^2-risk function and prove that this Galerkin estimator converges with the rate $T^{-\frac{s}{2s+3}}$ for functions g in the Sobolev space $H^s([-r, 0])$, $s > 0$. This rate is worse than those obtained in many classical cases. However, we prove a lower bound, stating that no estimator can attain a better rate of convergence in a minimax sense.

For discrete time observations of maximal distance Δ, the Galerkin estimator still attains the above asymptotic rate if Δ is roughly of order $T^{-1/2}$. In contrast, we prove that for observation intervals Δ, with Δ independent of T, the rate must deteriorate significantly by providing the rate estimate $T^{-\frac{s}{2s+6}}$ from below.

Furthermore, we construct an adaptive estimator by applying wavelet thresholding techniques to the corresponding ill-posed inverse problem. This nonlinear estimator attains the above minimax rate even for more general classes of Besov spaces $B^s_{p,\infty}$ with $p > \max(\frac{6}{2s+3}, 1)$. The restriction $p \geq \frac{6}{2s+3}$ is shown to hold for any estimator, hence to be inherently associated with the estimation problem.

Finally, a hypothesis test with a nonparametric alternative is constructed that could for instance serve to decide whether a trajectory has been generated by a stationary process with or without time delay. The test works for an L^2-separation rate between hypothesis and alternative of order $T^{-\frac{s}{2s+2.5}}$. This rate is again shown to be optimal among all conceivable tests.

For the proofs, the parameter dependence of the stationary solutions has to be studied in detail and the mapping properties of the associated covariance operators have to be determined exactly. Other results of general interest concern the mixing properties of the stationary solution, a case study for exponential weight functions and the approximation of the stationary process by discrete time autoregressive processes.

Contents

Chapter 1

Introduction

The classical research area of time series analysis is concerned with statistical inference for stochastic processes evolving in discrete time. The dynamics of these processes depend on their behaviour in the past and on some random perturbations. With the development of a powerful mathematical theory and the rise of computer power, continuous time stochastic processes gained a lot of attention in recent years. Many models are best set up in a continuous time setting so that the need of algorithms for similar tasks as for discrete time models arose. While the statistical theory for some classes of Markov processes, i.e. processes, where the dynamics only depend on the current state of the process and on random disturbances, is already far developed, the treatment of statistical questions for continuous time processes that depend also on the past is only just emerging.

Stochastic delay differential equations

Deterministic continuous time processes involving some time delay are widespread. Everybody will have experienced the problem of regulating the water temperature of a shower. Since turning the lever needs some time to result in a changed water temperature on the skin, we usually tend to adjust the temperature too strongly and an oscillation between higher and lower temperature than desired is the consequence. An easy mathematical "showering" model has been derived in Kolmanovskii and Myshkis (1992, Sec. 1.4):

$$T'(t) = -\kappa(T(t-h) - T_d),$$

where $T(t)$ denotes the water temperature at the mixer at time t, T_d the desired temperature value, h the time for the water to move from the mixer to the person's skin and $\kappa > 0$ is a constant modelling the reaction of the person to a wrong temperature. A phlegmatic person would choose a small value of κ and an energetic person a large value of κ. If we believe in this model, then mathematics can prove that a well-balanced character reacts optimally, because for too small κ the temperature adjusts only very slowly and for too large κ oscillations occur, maybe even with increasing amplitude leading to burns and frostbite.

There are also many examples of delay equations beyond the area of control theory. In population models the time delay is due to some pregnancy period or an age-dependent birth and death rate. The incubation period leads to time delays in epidemiology. The whole discussion, whether cyclic or anti-cyclic public spending better supports a steady economic growth, is basically a question of the time lag involved. A nice selection of continuous time models with time delay is presented in the book of Kolmanovskii and Myshkis (1992). In the literature many different names for the underlying equations are used: delay differential equations, functional differential equations, retarded differential equations, differential equations with aftereffect, with memory, with time lag etc.

Randomness plays a significant role, when the environment, in which the process evolves, is too complex to be modelled deterministically. There are plenty of possibilities to create differential equations involving time delay and randomness. When the randomness in a delay differential equation is due to some driving stochastic processes like for stochastic ordinary differential equations, then these equations are called stochastic delay differential equations (SDDEs). First results on SDDEs go back to Itô and Nisio (1964); the fundamental theory is presented in the monographs by Mohammed (1984) and Mao (1997). We shall here consider the so-called affine stochastic delay differential equations

$$\begin{aligned} dX(t) &= \left(\int_{-r}^{0} X(t+s)\,da(s)\right)dt + \sigma\,dW(t), && t \geq 0,\\ X(t) &= F(t), && t \in [-r, 0]. \end{aligned}$$

This is a usual stochastic differential equation with a one-dimensional Brownian motion $(W(t),\, t \geq 0)$, but involving a time delay in the drift. The drift is a mean over the past trajectory $(X(u),\, t-r \leq u \leq t)$ weighted by the signed measure a. Particular cases of the drift are point delays $\alpha_1 X(t-r_1) + \dots + \alpha_n X(t-r_n)$ with $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ and $r_1, \dots, r_n \geq 0$ or distributed delays $\int_{-r}^{0} X(t+s)g(s)ds$ with a Lebesgue-integrable function $g : [-r, 0] \to \mathbb{R}$. We assume $\sigma > 0$ and $r \geq 0$; the case of unbounded memory $r = \infty$ is not considered. In contrast to ordinary differential equations, we have to prescribe an initial function F on the interval $[-r, 0]$. Precise definitions of all the quantities involved will be given later.

The affine SDDEs form a basic class of autonomous stochastic delay differential equations, since they generalize linear deterministic delay differential equations as well as linear stochastic ordinary differential equations. The asymptotic behaviour of their solution processes is well understood even for more general driving processes than Brownian motion Mohammed and Scheutzow (1990) so that an investigation of statistical questions is feasible. The results obtained for this prototype model of a stochastic delay differential equation should be regarded as an indicator for the possibilities and the problems of statistical methods for other stochastic delay differential equations.

The estimation problem

Our goal here is the nonparametric estimation of the weight function in the case of distributed delay. We suppose that one trajectory $(X(t),\, -r \leq t \leq T)$ of a solution process is observed up to time $T > 0$ and try to estimate the weight function $g \in L^2([-r, 0])$, where the underlying weight a satisfies $da(u) = g(u)du$. As usual in nonparametric estimation theory, we shall investigate the asymptotic behaviour of our estimator. We consider the estimator for an observation period tending to infinity ($T \to \infty$). Another natural choice would be to consider small noise asymptotics ($\sigma \to 0$), but in this case the influence of the initial function becomes very important and the weight measure is not for all initial functions identifiable, e.g. not for $F = 0$. The solution process X is assumed to be stationary, which is not a severe restriction for weight measures a in the set M^-, to be introduced later. For other weight measures a the solution X is unstable in the long time limit, which is not very suitable for long time asymptotics.

Under the assumption of a parametric family of weight measures a in the point delay case, estimation results have been obtained by Mensch (1989), Kutoyants et al. (1992), Gushchin and Küchler (1999), Küchler and Kutoyants (2000) and Putschke (2001). The estimation problem is presented in the context of exponential families in the monograph of Küchler and Sørensen (1997). In the setting discussed above, i.e. $T \to \infty$ and stationarity, one obtains the usual $T^{-1/2}$-consistency and LAN-property for estimating the coefficients of the point measure a. Other parametric results concern the case of non-stationarity or the case that the position rather than the size of the point measure has to be estimated. The first to consider nonparametric estimation for affine SDDEs was Rothkirch (1993). He recognized the link with an ill-posed problem and derived a weakly consistent estimator. Kutoyants and Mourid (1994) proposed a nonparametric estimator of the deterministic trajectory for the small noise asymptotics, but did not address the problem of determining the weight measure a from this deterministic trajectory. The analytical problem, whether the weight measure is uniquely determined by a segment of a deterministic trajectory, has been considered by Verduyn Lunel (2000), but the partly positive results he obtains are non-constructive. It would be very interesting to see whether the last two approaches can be combined to yield an estimator in the small noise asymptotics for certain classes of initial functions.

The main results

In nonparametric statistics rates of convergence can usually only be proved for certain compact subclasses of the function family considered. Therefore, weight functions g that are s-times differentiable in an L^2-sense, that is g is in the Sobolev space $H^s([-r,0])$, are considered for $s > 0$. Choosing as the loss function the norm in $L^2([-r,0])$, we study for estimators $\hat{g}$ the uniform risk

$$\sup_{\|g\|_{H^s} \le S,\, v_0(g) \le -\delta} \mathbb{E}_g[\|\hat{g} - g\|_{L^2}^2]^{1/2}$$

over weight functions that are bounded in H^s-norm by some constant $S > 0$ and satisfy some later specified uniform stationarity assumption $v_0(g) \le -\delta < 0$. The symbol $\mathbb{E}_g$ denotes the expected value under the law of the stationary solution corresponding to the weight function g. The first main result is that the minimax rate of convergence for this risk is $T^{-\frac{s}{2s+3}}$ as T tends to infinity. This means that no estimator can attain a better rate and that we can construct an estimator with exactly this rate of convergence. In this sense, our proposed estimator is rate-optimal. This estimator is easy to implement numerically by the Galerkin projection method.

It is further proved that this so-called Galerkin estimator still attains this optimal rate, when the trajectory is only observed at discrete times of maximal distance Δ and the distance Δ is roughly of order $T^{-1/2}$ for $T \to \infty$. On the other hand, for equispaced observations of distance Δ, independent of T, it is shown that the rate cannot be better than $T^{-\frac{s}{2s+6}}$, a significant deterioration. The Galerkin estimator also exhibits a certain robustness with respect to the misspecification that the underlying true weight measure does not possess an L^2-density.

Inspired from classical nonparametric estimation theory, we show that the rate $T^{-\frac{s}{2s+3}}$ even holds for a larger subclass of weight functions when nonlinear methods are applied. By using a wavelet thresholding estimator we prove that this rate up to logarithmic factors is attainable for weight functions in the Besov space $B^s_{p,\infty}$ for $p > \max(\frac{6}{2s+3}, 1)$. This means that s derivatives in an L^p-sense suffice even for certain values of $p < 2$. On the other hand, the restriction $p \ge \frac{6}{2s+3}$ necessarily holds for any estimator in a minimax-sense, which follows from a corresponding risk lower bound. Moreover, the wavelet thresholding estimator is adaptive and also works for L^q-loss functions for $q \ne 2$. The only drawback is that its numerical calculation requires more effort than for the Galerkin estimator.

Finally, the problem of hypothesis testing with a nonparametric alternative is treated. The hypothesis is that a certain weight measure a_0, which is a linear combination of a measure with Lebesgue density and a point measure at the boundary points $-r$ and 0, is the true parameter. The alternative consists of all weight measures of similar linear combinations with s-regular density, which have a generalized L^2-distance $\rho > 0$ from a_0. For prescribed errors of the first and of the second kind we ask for the minimal value of ρ such that a test satisfies these error bounds. It turns out that in a similar minimax sense as above the best achievable rate for $\rho = \rho(T)$ is $T^{-\frac{s}{2s+2.5}}$ as T tends to infinity. This is better than the rate obtained from just using an estimator to derive the test statistics. The proposed test is constructive and easy to implement. For instance, it could be used to test whether some observed data comes from an Ornstein-Uhlenbeck process without delay or rather from an affine SDDE with proper delay.

All these results are new and somewhat surprising, because for many classical problems in nonparametric statistics like density estimation, regression or signal detection the minimax rates are better, i.e. $T^{-\frac{s}{2s+1}}$ for estimation and $T^{-\frac{s}{2s+0.5}}$ for testing, see e.g. Ibragimov and Khas'minskii (1981), Härdle et al. (1998) and Ingster (1993a). Observe the formal similarity of our estimation problem with the problem of estimating the drift function of a one-dimensional ergodic diffusion. Also in this case the better rate $T^{-\frac{s}{2s+1}}$ is obtained, consult Hoffmann (1999) for estimation results for diffusions. The reason why even the best estimator cannot attain this rate for the estimation of the weight function is due to the fact that the estimation problem is closely related to an ill-posed inverse problem. Heuristically, this may be explained by the fact that an integral over the weight function enters into the drift so that some sort of differentiation must be applied to recover the weight function. During the course of the proofs, it is however rather recommendable to keep a Gaussian signal plus noise model in mind: For a known covariance operator Q on $L^2([-r, 0])$ and centred Gaussian noise Γ on $L^2([-r, 0])$ with covariance operator Q the goal is to estimate the function $g \in L^2([-r, 0])$ from the observation of

$$Qg + T^{-1/2}\Gamma, \quad T > 0.$$

In the framework of Nussbaum and Pereverzev (1999), the covariance operator Q will be shown to have a degree of ill-posedness $\alpha = 2$ and the noise to have regularity $\rho = \frac{1}{2}$, whence for this model problem the minimax estimation rate is $T^{-\frac{s}{2s+2\alpha-2\rho}} = T^{-\frac{2s}{2s+3}}$ for $T \to \infty$.

Several different fields of mathematics are used for the derivation of the results. The analytical tools range from the theory of linear delay equations and abstract functional analysis via complex and Fourier analysis to function spaces and wavelets. The probabilistic methods rely on the theory of stochastic differential equations, Gaussian processes and Gaussian measures on Banach spaces. Results on the Galerkin scheme of numerical analysis are fundamental for the Galerkin estimator. The lower bound proofs for the minimax risk, the wavelet thresholding estimator and the results for the test follow mainly standard nonparametric statistical methods. The semigroup description of delay equations is only touched, because everything can be presented rather explicitly.

Some topics seem to be of general interest in the different fields. As far as we know, the Galerkin method with errors in the operator has until now been studied abstractly without any "real world" application in mind, which we provide. The stationary solutions of affine SDDEs form a subclass of stationary Gaussian processes, which have laws that are locally equivalent to the Wiener measure. From the highly developed theory of Gaussian processes even improvements of results for deterministic delay equations follow (cf. Remark 3). The methods of nonparametric statistics and those for ill-posed inverse problems are mostly designed for operators and for noise which satisfy very idealized conditions, so that it is remarkable that the methods can also be applied to less ideal situations. A beautiful example of the interplay between analysis and stochastics is given by the derivation of properties of the covariance operators: The injectivity follows from stochastic arguments using a change of measure, while the range is determined analytically by Fredholm theory. Thus, from the perspective of abstract mathematics nice applications of the theory and links between different branches are the delightful parts, while from an applied point of view the model assumptions are still rather idealized, but provide a reference for the investigation of more complicated models.

A general guideline

To grasp the main ideas, we shall now indicate the interdependency between the results, presented in the different sections. In order to derive uniform risk estimates, the dependence of the affine SDDE on the weight measure has to be described precisely (Sections 2.1, 2.3). The covariance operator Q_a corresponding to the stationary solution of the affine SDDE with weight measure a is studied on the space of L^2-functions (Sections 3.1, 3.2). The two main quantities for the estimation, the functions q_T and b_T are introduced and shown to converge to the covariance function q_a and a function b_a, respectively, in some Hölder space-norm with a rate $T^{-1/2}$ for $T \to \infty$ (Chapter 4). From the relation $b_a = Q_a(a)$ the idea to use $Q_T^{-1} b_T$ as an estimator is derived, where Q_T is the integral operator with kernel q_T. Due to the illposedness of the covariance operator Q_a, the problem has to be regularised by solving it in some finite dimensional space V_n, using the Galerkin projection method. The results for the classical Galerkin method are adapted to the case of perturbations of the operator and then applied to yield the announced results for the Galerkin estimator (Chapter 5).

The investigation of the wavelet thresholding operator (Section 7.1) relies heavily on the mapping properties of the covariance operator with respect to the scale of Besov spaces (Section 3.3). The main idea is to smooth b_T adaptively and to invert the operator equation afterwards. For the construction of the test (Sections 8.1, 8.2) the only new prerequisite is the β-mixing property of the stationary solution (Section 2.3), which is used for the estimates of some covariance terms. This mixing property is also of interest on its own. All the lower bound proofs for continuous time observations (Sections 6.1, 7.2, 8.3) rely on the likelihood ratio, which turns out to be a rather delicate object (Section 2.2).

Further results on affine SDDEs, which may stand on their own, concern the set of measures that admit stationary solutions, in particular topological properties of this set (Section 2.1) and its shape for a two-parameter family of exponential weight functions (Section 2.4), and the relationship between affine SDDEs and autoregressive schemes (Section 2.5). This last result supports the view of stochastic delay differential equations as continuous time analogues of time series models. We conclude by an outlook on some open problems and on variations of the model assumptions (Chapter 9).

Only the results that have not been known before have been listed and will be proved. Of course, the proofs sometimes follow ideas by other authors, to which then always credit is given. Whenever it was possible, results for general weight measures have been proved even if it had sufficed to consider absolutely continuous weight measures, in order to facilitate subsequent work on this subject. Some parts of Sections 2.1-2.3 and the appendix on function spaces and wavelets are written in a more colloquial style, when well-known results from the literature are discussed. The numbering always starts with the number of the chapter and of the section, the regularly used symbols are listed after this introduction.

Acknowledgement

During the work on this thesis I received very generous support from many people. Let me name a few of them in the framework of some subjective problem history.

Prof. Küchler (May '99) posed me the problem of nonparametric estimation for affine SDDEs, which turned out to provide a rich mathematical structure. Very soon, I became fascinated by this subject. Dr. Plato and Prof. Hämarik (autumn '99) gave me advice on the topic of ill-posed inverse problems; the results of Prof. Hämarik turned out to provide the key ideas for the Galerkin estimator in Section 5.1. Dr. Hoffmann (January '00) came to Berlin and sacrificed a whole week working with me on the lower bound problem. At the end of the week we have both got convinced that my upper bound $T^{-\frac{s}{2s+4}}$ and his lower bound $T^{-\frac{s}{2s+1}}$ had not been optimal; the lower bound proof of Section 6.1 owes him a lot. Prof. Behrends (spring '00) gave a proof of Lemma 9, on which the one presented here mainly relies. His research group provided me a very valuable forum for discussing my problems and ideas.

Prof. Jacod and Prof. Küchler made a stay possible at the Laboratoire de probabilités et modèles aléatoires in Paris (October '00 until March '01), which was financed by the EU-network "Dynstoch". Dr. Hoffmann created a physically (!) and scientifically challenging environment for me at Paris without neglecting social events. By working with Prof. Cohen and Dr. Hoffmann on the numerical side of the adaptive estimation problem, I also got a much better theoretical understanding. The work with Dr. Gobet and Dr. Hoffmann on lower bound problems for diffusions was similarly rewarding.

Prof. Werner helped me a lot by providing an excellent book on functional analysis, which unfortunately has not yet been translated into English, and by answering many of my analytical questions, including the silly ones. Prof. Küchler supervised and encouraged me throughout the years with a lot of personal effort so that I could start writing down my results (spring '01) earlier than expected. Evelyn Buckwar and Markus Riedle took on the burden of proof-reading the manuscript (autumn '01), the remaining mistakes being my responsibility alone. Working together with Markus Riedle in one office often saved me from attacks of mathematical frustration. Mathematical discussions and social events were equally enjoyable with the other, partly former, members of the research group: Eva, Ulrike, Bo and Hagen.

I gratefully thank all of them and the numerous other people supporting me consciously or unconsciously. Sometimes, the emotional support by my friends and parents has been more important for the progress than any scientific advice. This thesis would not have come into being without Astrid's joie de vivre and her amazing love.

Notation

The notation follows the usual conventions, nevertheless the general mathematical symbols that will be used are gathered in the first table. The notation of the different function spaces is presented in the second table (cf. also the appendix). The third table summarizes some frequently used quantities from the text.

General symbols

$A := B$	A is defined by B
$[a, b]$, (a, b)	closed, open interval from a to b
$\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{Z}$	$\{1, 2, \ldots\}$, $\{0, 1, \ldots\}$, $\{0, +1, -1, +2, -2, \ldots\}$
$\mathbb{R}$, $\mathbb{R}^+$, $\mathbb{R}^-$, $\mathbb{C}$	$(-\infty, \infty)$, $[0, \infty)$, $(-\infty, 0]$, complex numbers
$\mathrm{Re}(z), \mathrm{Im}(z), \bar{z}$	real part, imaginary part, complex conjugate of $z \in \mathbb{C}$
$\lfloor x \rfloor$	largest integer smaller or equal to $x \in \mathbb{R}$
$\lceil x \rceil$	smallest integer larger or equal to $x \in \mathbb{R}$
$a \vee b$, $a \wedge b$	maximum, minimum of a and b
$a \ll b$, $a \approx b$	a is much smaller than, approximately equal to b
$\lvert x \rvert$	modulus of $x \in \mathbb{R}$ or Euclidean norm of $x \in \mathbb{R}^d$
$\lvert S \rvert$, S a set	cardinality of S
$A \subset B$	A is contained in B or $A = B$
∂S, $S \subset \mathbb{C}$	boundary of S in $\mathbb{C}$
$\mathrm{span}(v, w, \ldots)$	the subspace spanned by $v, w, \ldots$
$U + V$, $U \oplus V$	the sum, the direct sum ($U \cap V = \{0\}$) of U and V
$\dim V$, $\operatorname{codim} V$	linear dimension, codimension of V
$\operatorname{ran} T$, $\ker T$	range and kernel of the operator T
Id	identity operator or matrix
$\det(M)$	determinant of M
$\lVert T \rVert$, $\lVert T \rVert_{X \to Y}$	operator norm of $T : X \to Y$
$\lVert T \rVert_{HS}$	Hilbert-Schmidt norm of T
$\bullet$	multiplication sign
$f(\bullet)$, $g(\bullet_1, \bullet_2)$	the functions $x \mapsto f(x)$, $(x_1, x_2) \mapsto g(x_1, x_2)$
$\mathrm{supp}(f)$	support of the function or distribution f
$f\vert_S$	function f restricted to the set S
$f', f'', f^{(m)}$	first, second, m-fold (weak) derivative of f
$f'(a+)$	derivative of f at a to the right
$\mathbf{1}_S$	indicator function of the set S
$\hat{f}$, $\mathcal{F}(f)$	$\hat{f}(\xi) = \mathcal{F}(f)(\xi) = \int_{\mathbb{R}} f(t) e^{-i\xi t} dt$ or estimator $\hat{f}$ of f
$\hat{a}$, $\mathcal{F}(a)$, $a \in M(I)$	$\hat{a}(\xi) = \mathcal{F}(a)(\xi) = \int_I e^{-i\xi t} da(t)$
log	natural logarithm
cos, sin, cosh, sinh	(hyperbolic) trigonometric functions
$\mathbb{P}$, $\mathbb{E}$, Var, Cov	probability, expected value, variance and covariance
$\mathbb{E}_g$, Var_g, Cov_g	as before with respect to the probability measure $\mathbb{P}_g$
$\mathcal{L}(X)$, $X \sim \mathbb{P}$	the law of X, $\mathcal{L}(X) = \mathbb{P}$
$\sigma(Z_i, i \in I)$	σ-algebra generated by $(Z_i)_{i \in I}$

δ_x	Dirac measure at x
$A \lesssim B$	$A = O(B)$, i.e. $\exists\, c > 0\, \forall p:\ A(p) \le cB(p)$ (p parameter)
$A \gtrsim B$	$B \lesssim A$
$A \sim B$	$A \lesssim B$ and $B \lesssim A$

Function spaces and norms

$L^p(I)$	p-integrable functions f on I: $\int_I \lvert f\rvert^p < \infty$
$H^s(I)$	L^2-Sobolev space of regularity s on I
$\lVert\bullet\rVert_s$	norm in H^s
$B^s_{p,\alpha}(I)$	Besov space on I
$\lVert\bullet\rVert_{s,p,\alpha}$	norm in $B^s_{p,\alpha}$ or in $\mathcal{W}^s_{p,\alpha}$
$\langle \bullet_1, \bullet_2\rangle$, $\lVert\bullet\rVert$	scalar product in L^2 or dual pairing, L^2-norm
$C(I) = C^0(I)$	$\{f : I \to \mathbb{R} \mid f \text{ continuous}\}$
$C_{\mathbb{C}}(I)$	$\{f : I \to \mathbb{C} \mid f \text{ continuous}\}$
$\lVert f\rVert_\infty$	$\sup_x \lvert f(x)\rvert$
$C^\alpha(I)$, $0 < \alpha < 1$	α-Hölder continuous functions for $0 < \alpha < 1$
$C^s(I)$, $s = m + \alpha$	functions f with $f^{(i)} \in C^\alpha$, $\forall\, i \le m \in \mathbb{N}$; $0 \le \alpha < 1$
$C^{m,1}(I)$	functions $f \in C^m(I)$ with $f^{(m)}$ Lipschitz continuous
$\mathcal{D}(\mathbb{R}^+)$	Skorohod space on $\mathbb{R}^+$
$\mathcal{H}^2$	Hardy space of holomorphic functions in the upper half plane
$M(I)$	space of finite signed measures on I
$\lVert\mu\rVert_{TV}$	total variation norm: $\lVert\mu\rVert_{TV} = \sup_{\lVert f\rVert_\infty = 1} \int f\, d\mu$

Specific definitions

$W(t)$	standard one-dimensional Brownian motion at time t
$X(t)$, $X^{(a)}(t)$	solution process to SDDE (with weight a) at time t
$(\mathcal{F}_t)$; $\mathcal{F}^X_T$	filtration for $(W(t), t \ge 0)$; $\sigma(X(t), 0 \le t \le T)$
a, g	weight measure, weight function of SDDE
r	length of memory in SDDE
χ, χ_a	characteristic function of delay equation (with weight a)
v_0, $v_0(a)$	largest real part of zeros of χ, χ_a
$M^- = M^-([-r,0])$	$\{\mu \in M([-r,0]) \mid v_0(\mu) < 0\}$
$M(R,\delta)$	$\{\mu \in M([-r,0]) \mid \lVert\mu\rVert_{TV} \le R,\ v_0(\mu) \le -\delta\}$
$M(s,S,\delta)$	$\{f \in H^s([-r,0]) \mid \lVert f\rVert_s \le S,\ v_0(f) \le -\delta\}$
$M(s,p,S,\delta)$	$\{\mu \in \mathcal{W}^s_{p,\infty} \mid \lVert\mu\rVert_{s,p,\infty} \le S,\ v_0(\mu) \le -\delta\}$
$\mu_n \xrightarrow{w} \mu$	stochastic weak convergence: $\int f\, d\mu_n \to \int f\, d\mu\ \forall f \in C$
Δ	maximal distance between discrete time points
$\mathcal{W}^s_{p,\alpha}$	$\mathcal{W}^s_{p,\alpha} \cong B^s_{p,\alpha}([-r,0]) \oplus \operatorname{span}(\delta_{-r}, \delta_0)$
q, q_a	(auto-)covariance function of SDDE (with weight a)
Q, Q_a	covariance operator on $[-r,0]$ with kernel q, q_a
$\mathcal{Q}$	covariance matrix $q((i-j)\Delta)_{0 \le i,j \le N}$
q_T, Q_T	$q_T(u,v) = \int_0^T X(t+u)X(t+v)\, dt$, integral operator
b_T	$b_T(s) = \int_0^T X(t+s)\, dX(t)$
$\hat{y}_T$	adaptive estimate of $Q_a a$
$\hat{q}_T$, $\widehat{Q}_T$	estimate of q_a based on $\hat{y}_T$, integral operator
$\hat{a}_T$	adaptive estimator of $a \in \mathcal{W}^s_{p,\infty}$
V_n	approximation space
P_n	L^2-orthogonal projection on V_n
$\hat{g}_{T,n}$	Galerkin estimator of g
λ, $\lvert\lambda\rvert$	multi-index $\lambda = (j,k)$, $\lvert\lambda\rvert = j$
ψ, ψ_{jk}, ψ_λ	wavelet, $\psi_\lambda(x) = \psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$

$\Lambda_T(X^{(1)}, X^{(2)})$	Radon-Nikodym derivative $\frac{\mathcal{L}(X^{(1)})}{\mathcal{L}(X^{(2)})}$ on $C([0,T])$
$\mathcal{H}_0 = \{a_0\}$	test hypothesis consisting of the weight a_0
$\mathcal{K}_\rho$	test alternative separated from $\mathcal{H}_0$ by a ball of radius ρ
$\mathcal{T}$; φ, Φ	main test statistic; decision functions

Chapter 2

Affine stochastic delay equations

The results of this chapter on stochastic delay differential equations prepare the grounds for the subsequent statistical applications. We start with a brief review of the deterministic theory of linear autonomous delay equations. For future estimation purposes it is shown that in the right topology the characteristic functions depend continuously on the weights. Moreover, the set of weights yielding exponentially stable solutions is an open subset in the space of signed measures. Basic facts about affine SDDEs are contained in the second section, in particular an expression for the Radon-Nikodym derivative or likelihood function is presented. In the fundamental third section on stationary solutions, results on the existence of stationary solutions and on their spectral density are used to investigate the mixing properties of these solutions and the parameter dependence of the covariance functions. As an example, SDDEs with an exponential weight function are considered in the fourth section. In this case the region of stability and the covariance function can be determined quite explicitly. In the last section it is shown that the stationary solution of an affine SDDE can be approximated in the Skorohod topology by piecewise constant stationary autoregressive processes. This supports the view of affine SDDEs as continuous time analogues of autoregressive schemes.

2.1 The deterministic linear theory

The theory of deterministic delay equations that are linear, autonomous and have a bounded memory is well understood. The notions and results without proof that are presented in this section can be found for instance in Chapter 7 of Hale and Verduyn Lunel (1993) and Chapters I-IV of Diekmann et al. (1995), to which we refer for exact statements and proofs. The only deviation from the classical terminology will be the (equivalent) use of finite signed measures instead of normalized bounded variation (NBV) functions. New deterministic results mainly concern the parameter dependence of certain quantities, which is vital for the establishment of minimax results in the subsequent statistical applications.

Let $a \in M([-r, 0])$ be a signed measure ("the weight") on $[-r, 0]$ with finite total variation norm $\|a\|_{TV}$ for $r \in [0, \infty)$ ("the length of memory") and let $F : [-r, 0] \to \mathbb{R}$ be a deterministic initial function. If not stated differently, we shall henceforth assume $r > 0$ to avoid trivialities. Then a deterministic, one-dimensional, linear and autonomous delay equation with bounded memory can be written in the form:

$$\begin{aligned} x'(t) &= \int_{-r}^{0} x(t+s)\,da(s), && t \geq 0, \\ x(t) &= F(t), && t \in [-r,0]. \end{aligned} \tag{2.1.1}$$

By $\int_{-r}^{0}$ we always mean the integral over the closed interval $[-r,0]$. Mostly, F is assumed to be continuous, i.e. $F \in C([-r,0])$. Then, by the Riesz representation theorem, the differential equation (2.1.1) is the general form of an autonomous first order equation where the derivative $x'(t)$ is a continuous linear functional of the function segment $(x(t+s),\, s \in [-r,0])$ regarded as an element of $C([-r,0])$. The dynamics of x may be described by saying that the growth at time t equals a weighted mean of the values between $t-r$ and t. The case $a = \alpha\delta_0$ corresponds to the ordinary differential equation $x'(t) = \alpha x(t)$.

It is possible to construct a fundamental solution and a characteristic equation for (2.1.1) which are of the same importance in the representation and analysis of solutions as in the case of ordinary differential equations.

Definition 1. *A function x_0 that is absolutely continuous on $[0,\infty)$ and solves (2.1.1) in a weak sense with initial values $x_0(0) = 1$ and $x_0(t) = 0$ for $t \in [-r,0)$ is called* fundamental solution *of* (2.1.1).

The function $\chi : \mathbb{C} \to \mathbb{C}$ defined by

$$\chi(\lambda) := \chi_a(\lambda) := \lambda - \int_{-r}^{0} e^{\lambda s} da(s), \tag{2.1.2}$$

is called characteristic function *associated to equation* (2.1.1) *or just to the weight a. We set*

$$v_0 := v_0(a) := \sup\{\mathrm{Re}(\lambda) \mid \chi_a(\lambda) = 0\}. \tag{2.1.3}$$

The set of all $a \in M([-r,0])$ with $v_0(a) < 0$ will be denoted by $M^-([-r,0])$ or just M^-. Its elements will be called M^--weights. *Furthermore, for $R > 0$, $\delta \in \mathbb{R}$ we set*

$$M(R,\delta) := \{a \in M([-r,0]) \mid \|a\|_{TV} \leq R,\ v_0(a) \leq -\delta\}. \tag{2.1.4}$$

A standard result is that a solution $x \in C([0,\infty))$ to (2.1.1) with continuous initial function F exists and is unique. Existence and uniqueness of the fundamental solution x_0 also holds true. Furthermore, such a solution x satisfies

$$x(t) = F(0)x_0(t) + \int_{-r}^{0}\int_{-s}^{0} x_0(t+s-u)F(u)\,du\,da(s), \quad t \geq 0. \tag{2.1.5}$$

The set $\{\lambda \in \mathbb{C} \mid \mathrm{Re}(\lambda) \geq \beta,\ \chi(\lambda) = 0\}$ is finite for all $\beta \in \mathbb{R}$ due to the holomorphy of χ and to the estimate

$$|\chi(\lambda)| \geq |\lambda| - \|a\|_{TV} e^{|\beta| r} > 0, \quad \text{for } \mathrm{Re}(\lambda) \geq \beta,\ |\lambda| > \|a\|_{TV} e^{|\beta| r}. \tag{2.1.6}$$

Consequently, the value $v_0(a)$ is always finite and not larger than $\|a\|_{TV}$. If $\chi(\lambda) = 0$ holds for some $\lambda \in \mathbb{C}$, then $\mathrm{Re}(e^{\lambda t})$ and $\mathrm{Im}(e^{\lambda t})$ are solutions of the delay equation (2.1.1) on the whole real line. For all continuous initial functions F the asymptotic growth of $x(t)$ for $t \to \infty$ can be estimated by

$$|x(t)| \lesssim e^{vt}, \quad t \geq 0, \tag{2.1.7}$$

for any $v > v_0$, i.e. there is a constant $C = C(v,F)$ with $|x(t)| \leq Ce^{vt}$ for all $t \geq 0$. Thus, if a is an M^--weight, then the solutions of (2.1.1) converge to zero exponentially fast. This asymptotic behaviour also extends to the derivatives:

$$|x'(t)| = \left|\int_{-r}^{0} x(t+s)\,da(s)\right| \leq \sup_{t-r \leq s \leq t} |x(s)| \|a\|_{TV} \lesssim e^{vt}, \quad t \geq 0. \tag{2.1.8}$$

The fundamental solution and the characteristic function are linked by the Laplace transform:

$$\int_0^\infty e^{-\lambda t} x_0(t)\,dt = \frac{1}{\chi(\lambda)} \text{ for all } \lambda \in \mathbb{C} \text{ with } \mathrm{Re}(\lambda) > v_0. \tag{2.1.9}$$

Example 1. *The linear ordinary differential equation $x'(t) = \alpha x(t)$ with initial condition $x(0) = F(0)$ is a particular instance of a "delay equation" in the above sense with $r = 0$ and $a = \alpha\delta_0$. Its fundamental solution is $x_0(t) = e^{\alpha t}$ and its characteristic function is $\chi(\lambda) = \lambda - \alpha$ with $v_0 = \alpha$. In this case, (2.1.5), (2.1.7) and (2.1.9) can easily be checked. We shall frequently refer to this degenerate case because then the subsequent calculations will always be explicit.*

Since the characteristic function is nonlinear, the set $M^-([-r,0])$ is difficult to determine (cf. Section 2.4), but there are at least some simple conditions on a for $v_0(a) < 0$ and $v_0(a) > 0$.

Lemma 1.

1. *If $-a \in M([-r,0])$ is a positive measure with $\|a\|_{TV} = -a([-r,0]) \in (0, \frac{\pi}{2r})$, then $v_0(a) < 0$ holds.*

2. *For weights a with $a([-r,0]) > 0$ always $v_0(a) > 0$ holds.*

Proof.

1. Let us consider $x, y \in \mathbb{R}$ with $\chi_a(x+iy) = 0$. By looking at real and imaginary part separately, we find

$$x = \int_{-r}^0 e^{xs} \cos(ys)\,da(s), \qquad y = \int_{-r}^0 e^{xs} \sin(ys)\,da(s).$$

   Suppose $x \geq 0$. Then from the second identity the inequality

$$|y| \leq \|e^{x\bullet} \sin(y\bullet)\|_\infty \|a\|_{TV} < \tfrac{\pi}{2r}$$

   follows. Inserting this into the first identity shows that the integrand there is strictly positive, thus the contradiction $x < 0$ follows.

2. For $a([-r,0]) > 0$ the characteristic function χ_a has always a positive real root due to

$$\chi_a(0) = -a([-r,0]) < 0 \text{ and } \chi_a(x) \geq x - \|a\|_{TV} \to \infty \text{ for } x \to \infty.$$

□

The bound $\frac{\pi}{2r}$ for the first result is sharp as the example $a = -\frac{\pi}{2r}\delta_{-r}$ with $\chi_a(\pm\frac{\pi}{2r}i) = 0$ shows Küchler and Mensch (1992, Cor. 2.9). One can show that if the negative feedback induced by a is too strong, then oscillations with exponentially growing amplitude may occur. The second result states that in the mean positive weights always allow for exploding solutions.

It will turn out that the right topology to consider on $M([-r,0])$ is the weak* topology in functional analytic language or the weak topology in probabilistic terms. We shall henceforth use the second terminology. For signed measures this topology is not metrizable, though the relative topology on the (norm) unit ball is metrizable owing to the separability of $C([-r,0])$ Rudin (1991, Thm. 3.16). Moreover, any ball in $M([-r,0])$ is weakly compact by the Banach-Alaoglu theorem Rudin (1991, Thm. 3.15) and any weakly convergent sequence is norm bounded by the uniform boundedness principle Rudin (1991, Thm. 2.5). Compactness will prove to be essential so that we shall only deal with weakly converging sequences and not with nets or filters.

Definition 2. *We say that a sequence* $(a_n)_{n\in\mathbb{N}} \subset M([-r,0])$ converges weakly *to* $a \in M([-r,0])$, *if*

$$\lim_{n\to\infty} \int_{-r}^{0} f(s)\, da_n(s) = \int_{-r}^{0} f(s)\, da(s) \qquad \textit{for all } f \in C([-r,0])$$

holds, and write $a_n \xrightarrow{w} a$. *If* $a_n \xrightarrow{w} a$ *implies* $f(a_n) \to f(a)$ *for some function* f *with values in a metric space, we say that* f *is* weakly continuous, *instead of using the more precise attribute weakly sequentially continuous.*

We prove a lemma about the convergence of characteristic functions and show two topological properties of the set M^-. Two-dimensional sections of M^- already reveal a complicated geometric structure, in particular M^- is not convex and its boundary is not differentiable (see Diekmann et al. (1995, Chap. 11) and the forthcoming Figures 2.4.2 and 2.4.3 on the pages 30 and 31).

Lemma 2. *If a sequence* $(a_n) \subset M([-r,0])$ *converges weakly to* $a \in M([-r,0])$, *then the characteristic functions* χ_{a_n} *converge uniformly on compact sets to the characteristic function* χ_a.

Proof. Let $K \subset \mathbb{C}$ be compact. The convergence $a_n \xrightarrow{w} a$ implies $\chi_{a_n}(z) \to \chi_a(z)$ pointwise for all $z \in \mathbb{C}$. Setting for $f \in C_{\mathbb{C}}([-r,0])$, the space of $\mathbb{C}$-valued continuous functions with supremum norm,

$$l_n(f) := \int_{-r}^{0} f\, d(a_n - a),$$

we obtain $\chi_{a_n}(\lambda) - \chi_a(\lambda) = l_n(e_\lambda)$, $e_\lambda(t) := e^{\lambda t}$. Now observe that the family $F = \{e_\lambda \mid \lambda \in K\}$ is a precompact subset of $C_{\mathbb{C}}([-r,0])$. Indeed, by compactness of K

$$\sup_{\lambda\in K} \|e'_\lambda\|_\infty = \sup_{(\lambda,s)\in K\times[-r,0]} |\lambda e^{\lambda s}| < \infty$$

holds so that F is equicontinuous and the Arzelà-Ascoli theorem Rudin (1991, Thm. A5) applies. Set $S := \sup_n \|a_n - a\|_{TV}$, which is finite by the argument given before Definition 2. Suppose that there exists an $\varepsilon > 0$ and a sequence $(\lambda_n)_n \subset K$ with $|l_n(e_{\lambda_n})| > \varepsilon$ for all n, i.e. that l_n does not converge uniformly to zero on F. We can then assume that $e_{\lambda_n} \to f$ holds for some $f \in C_{\mathbb{C}}([-r,0])$ by passing to a convergent subsequence. However, weak convergence then implies for $n \to \infty$

$$|l_n(e_{\lambda_n})| \le |l_n(e_{\lambda_n} - f)| + |l_n(f)| \le S\|e_{\lambda_n} - f\|_\infty + |l_n(f)| \to 0,$$

which contradicts the assumption. Therefore, l_n converges to zero uniformly on F and hence χ_{a_n} converges to χ_a uniformly on K. □

Theorem 1.

1. *The set* $M^-([-r,0])$ *is pathwise connected in total variation norm.*

2. *Weak convergence of* $(a_n) \subset M([-r,0])$ *to* $a \in M([-r,0])$ *implies* $v_0(a_n) \to v_0(a)$.

3. $M^-([-r,0])$ *is a weakly sequentially open subset of* $M([-r,0])$, *which means that for* $a \in M^-$ *and* $a_n \xrightarrow{w} a$ *the weights* a_n *lie in* $M^-([-r,0])$ *for sufficiently large values of* n.

Proof.

1. The key for proving connectedness is the following translation relation:

$$\chi_a(\lambda+\tau) = \lambda + \tau - \int_{-r}^{0} e^{\lambda s} e^{\tau s} da(s) = \chi_{a_\tau}(\lambda), \quad \lambda \in \mathbb{C},\ \tau \in \mathbb{R},$$

   with $da_\tau(s) = -\tau d\delta_0(s) + e^{\tau s} da(s)$. The mapping $\tau \mapsto a_\tau$ is obviously norm continuous. Given an M^--weight a, we shall construct a path to the M^--weight $-A\delta_0$ for any given $A > \|a\|_{TV}$. Then two given measures can be connected via $-A\delta_0$ for A large enough.

   Let a be a measure in M^- and $A > \|a\|_{TV}$. For $\tau \geq 0$ the measures a_τ remain in M^- by the translation relation. Therefore $a = a_0$ and a_A are connected in M^-. Now consider the path $h \mapsto c_h$ from $-A\delta_0$ to a_A with $dc_h(s) = -A\, d\delta_0(s) + he^{As} da(s)$ and $h \in [0,1]$. This path is again norm continuous and for $\mathrm{Re}(\lambda) \geq 0$ the estimate (2.1.6) yields

$$|\chi_{c_h}(\lambda)| \geq |\lambda + A| - \|c_h + A\delta_0\|_{TV} \geq A - \|a\|_{TV} > 0.$$

   Hence, there is a norm continuous path between a_A and $-A\delta_0$ which lies completely in M^- and the first statement has been proved.

2. Let $H_\rho := \{\lambda \in \mathbb{C} \mid \mathrm{Re}(\lambda) > \rho\}$ be a left-bounded half plane and for $\mu \in M([-r,0])$ let $N(\mu,\rho)$ be the number of zeros of χ_μ in H_ρ counted according to their multiplicity. For the second statement it suffices to show that $N(a_n,\rho)$ equals $N(a,\rho)$ for a dense set of real values ρ and $n = n(\rho)$ sufficiently large, since then $N(a_n, v_0(a)+\varepsilon) = 0$ and $N(a_n, v_0(a)-\varepsilon) \geq 1$ holds for any $\varepsilon > 0$ and large n.

   Choose $S > \sup_n \|a_n\|_{TV}$ and $\rho \in \mathbb{R}$ such that no zero of χ_a has real part ρ. Since χ_a has only countably many zeros, the set of admissible values of ρ is dense in $\mathbb{R}$. We consider the region

$$R := \{\lambda \in H_\rho \mid |\lambda| > Se^{|\rho| r}\}.$$

   By the estimate (2.1.6) we infer that there are no zeros of a_n or of a in the closure of R. Since $H_\rho \setminus R$ is a semi circle, the argument principle of complex analysis yields for a counterclockwise contour Γ of index 1 on the boundary $\partial(H_\rho \setminus R)$

$$N(a_n,\rho) = \frac{1}{2\pi i} \oint_\Gamma \frac{\chi'_{a_n}(z)}{\chi_{a_n}(z)}\, dz,$$

   if the integrand is finite Ahlfors (1979, Thm. 20). Because $\chi_{a_n} \to \chi_a$ holds uniformly in any compact neighbourhood of $\partial(H_\rho \setminus R)$ by Lemma 2, it is known from complex analysis that the integrand converges uniformly Ahlfors (1979, Thm. 13) and hence $N(a_n,\rho)$ eventually equals $N(a,\rho)$ for $n \to \infty$.

3. This is just the preceding statement applied to $a \in M^-$, i.e. $v_0(a) < 0$.

□

Remarks 1.

- *Since the norm topology is stronger than the weak topology on $M([-r,0])$, the set M^- is also pathwise connected in the weak topology and $\|a_n - a\|_{TV} \to 0$ also implies $\chi_{a_n} \to \chi_a$ uniformly on compact sets. In particular, M^- is an open connected subset of $M([-r,0])$ in the norm topology.*

- *The second statement of the theorem can easily be generalized. Suppose $a_n \xrightarrow{w} a$ holds and D is a domain, which is bounded to the left, has sufficiently smooth boundary and no zeros of χ_a on the boundary. Then the number of zeros of χ_{a_n} on D eventually, i.e. for large values of n, equals the number of zeros of a. The proof is achieved by applying the argument principle to χ_{a_n} on the set $D \setminus R$ with R as in the preceding proof. This property does not hold any more for general unbounded domains; for instance, $\chi_{\delta_{-r}}$ has infinitely many zeros in $\mathbb{C}$ for $r > 0$, whereas χ_{δ_0} has exactly one ("the zeros disappear at $-\infty$ as r tends to zero").*

Corollary 1. *For any $R > 0$ and $\delta \in \mathbb{R}$ the set of weights $M(R, \delta)$, defined in (2.1.4), is weakly compact in $M([-r, 0])$.*

Proof. As stated before Definition 2, the ball $\{a \mid \|a\|_{TV} \leq R\}$ is weakly compact and metrizable. Theorem 1 then shows that $a \mapsto v_0(a)$ is weakly continuous on this ball, whence $M(R, \delta)$ is a weakly closed subset of the weakly compact ball and thus itself weakly compact. □

2.2 The solution of an affine SDDE

Inhomogeneous linear or affine stochastic delay differential equations (affine SDDEs) result from the perturbation of a deterministic linear equation by additive white noise. More precisely, assume that $(W_t,\, t \geq 0)$ is a standard one-dimensional Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ and that the filtration $(\mathcal{F}_t)$ satisfies the usual conditions, concerning right-continuity and $\mathbb{P}$-negligible sets Karatzas and Shreve (1991, Def. 2.25). Given an $\mathcal{F}_0$-measurable and $\mathbb{P}$-almost surely continuous (random) initial function $F : [-r, 0] \to \mathbb{R}$ with $\mathbb{E}[\|F\|_\infty^2] < \infty$, the affine SDDE

$$\begin{aligned} dX(t) &= \left(\int_{-r}^{0} X(t+s)\, da(s) \right) dt + dW(t), && t \geq 0, \\ X(t) &= F(t), && t \in [-r, 0] \end{aligned} \tag{2.2.10}$$

has a unique strong solution $X = (X(t), t \geq -r)$ in $C([-r, \infty))$. Following Mao (1997), a process X on $(\Omega, \mathcal{F}, \mathbb{P})$ is called a strong solution if X is a.s.-continuous, $(\mathcal{F}_t)$-adapted, satisfies $X(t) = F(t)$ for $t \in [-r, 0]$ and

$$X(t) = F(0) + \int_0^t \int_{-r}^0 X(t+s)\, da(s)\, dt + W(t), \quad \forall\, t \geq 0.$$

The solution is called unique if any other solution X' is indistinguishable from X, i.e. satisfies $\mathbb{P}(X(t) = X'(t), \forall t \geq 0) = 1$. The existence and uniqueness result follows from Mao (1997, Thm. 5.2.2), because the linear drift functional and the constant diffusion coefficient both satisfy the uniform Lipschitz and the linear growth condition. If $\mathbb{E}[\|F\|_\infty^p] < \infty$ holds for $p \geq 2$, then the solution process also has uniform moments of order p Mao (1997, Thm. 5.4.1):

$$\mathbb{E}\left[\sup_{t \leq T} |X(t)|^p \right] < \infty \qquad \text{for all } T \geq 0. \tag{2.2.11}$$

Several remarks seem to be in order. Note that the drift functional is measurable and non-anticipative so that X belongs to the class of Itô processes, but it is not a diffusion type process unless F is $\sigma(X(0))$-measurable Liptser and Shiryaev (2001, Chap. 4.2). It is not a Markov process if the trivial case $\mathrm{supp}(a) \subset \{0\}$ is excluded; however, it may be lifted to a Markov process with state space $C([-r, 0])$

as presented in Mohammed (1984, Chapters III-IV). Unfortunately, the transition semigroup on $C_{ucb}(C([-r,0]))$, the uniformly continuous and bounded functions on $C([-r,0])$, is only weakly continuous and the weak generator has to be determined on this rather complicated space.

The space transformation $Y(t) := \sigma X(t)$, $\sigma > 0$, transforms (2.2.10) to the same equation in terms of Y, but with diffusion coefficient σ. Therefore and since the diffusion coefficient can be exactly determined from a continuous observation of a trajectory, $\sigma = 1$ will be assumed throughout this work. Mean inverting drift terms of the form $\int_{-r}^{0} X(t+s)da(s) + m$ with $m \in \mathbb{R}$ can also be reduced to $m = 0$ in the case $a([-r,0]) \neq 0$ by regarding $\tilde{X}(t) = X(t) + a([-r,0])^{-1}m$. By a time transformation one could even assume $r = 1$, but leaving r in a general form will not complicate the expressions.

As in the deterministic case, a variation of constants formula yields a representation of $X(t)$ for $t \geq 0$ in terms of the fundamental solution x_0, the initial function F and a stochastic integral with respect to W Mohammed et al. (1986):

$$X(t) = F(0)x_0(t) + \int_{-r}^{0}\int_{-s}^{0} x_0(t+s-u)F(u)\,du\,da(s) + \int_0^t x_0(t-s)\,dW(s). \quad (2.2.12)$$

Since x_0 is absolutely continuous for $t \geq 0$, it suffices to interpret the stochastic integral in the Wiener sense, i.e. as being defined by partial integration, which immediately yields the existence of a continuous version.

For statistical purposes the knowledge of the Radon-Nikodym derivative (likelihood function) of the considered processes is (almost) indispensable.

Definition 3. *Define the sub-σ-algebra $\mathcal{F}_T^X := \sigma(X(t),\, 0 \leq t \leq T)$ of $\mathcal{F}_T$.*

Theorem 2. *Let μ_X denote the law on $C([0,T])$ induced by the solution process $(X(t), 0 \leq t \leq T)$ of (2.2.10) with a continuous initial condition F satisfying for some $\varepsilon > 0$ the exponential moment condition $\mathbb{E}[\exp(\varepsilon\|F\|_\infty^2)] < \infty$. If μ_W denotes the law induced by $(X(0) + W(t), 0 \leq t \leq T)$ on $C([0,T])$, then μ_X and μ_W are mutually absolutely continuous with Radon-Nikodym derivative:*

$$\Lambda_T(X(0)+W, X) := \frac{d\mu_W}{d\mu_X} =$$
$$\mathbb{E}\left[\exp\left(-\int_0^T\int_{-r}^0 X(t+s)\,da(s)\,dX(t) + \frac{1}{2}\int_0^T\left(\int_{-r}^0 X(t+s)\,da(s)\right)^2 dt\right)\middle|\, \mathcal{F}_T^X\right]$$

Remark 1. *The statement is to be understood in the sense that the conditional expectation, which is an $\mathcal{F}_T^X$-measurable map on Ω, is μ_X-a.s. canonically identified with a Borel-measurable functional on $C([0,T])$ (cf. the discussion at the beginning of Section 7.1 in Liptser and Shiryaev (2001)).*

Proof. The proof is an application of Liptser and Shiryaev (2001, Thm. 7.1) regarding Itô processes, adapted to nonzero initial value (cf. their more general Thm. 7.18). It remains to verify the two properties

$$\mathbb{P}\left(\int_0^T\left(\int_{-r}^0 X(t+s)\,da(s)\right)^2 dt < \infty\right) = 1,$$
$$\mathbb{E}\left[\exp\left(-\int_0^T\int_{-r}^0 X(t+s)\,da(s)\,dX(t) - \frac{1}{2}\int_0^T\left(\int_{-r}^0 X(t+s)\,da(s)\right)^2 dt\right)\right] = 1.$$

The first property follows from the boundedness of X in (2.2.11). The second one can be deduced from the representation (2.2.12), which shows that $|X(t)|$ is bounded pathwise by $K_T(\|F\|_\infty + \sup_{0\le s\le T}|W(s)|)$ for $0 \le t \le T$, K_T a constant, since the stochastic integral may be calculated pathwise by partial integration. Following the proof of Karatzas and Shreve (1991, Cor. 3.5.16), we obtain for all $0 \le t_{n-1} < t_n \le T$ with $t_n - t_{n-1} < K_T^{-2}\min(\varepsilon, (2T)^{-1})$ by independence of F and W and by Doob's maximal inequality for submartingales

$$\begin{aligned}
&\mathbb{E}\Big[\exp\Big(\frac{1}{2}\int_{t_{n-1}}^{t_n} |X(s)|^2\,ds\Big)\Big] \\
&\qquad \le \mathbb{E}\Big[\exp\Big((t_n - t_{n-1})K_T^2(\|F\|_\infty^2 + \sup_{0\le s\le T} W(s)^2)\Big)\Big] \\
&\qquad = \mathbb{E}\Big[\exp\Big((t_n - t_{n-1})K_T^2\|F\|_\infty^2\Big)\Big]\,\mathbb{E}\Big[\exp\Big((t_n - t_{n-1})K_T^2 \sup_{0\le s\le T} W(s)^2\Big)\Big] \\
&\qquad \le \mathbb{E}\Big[\exp(\varepsilon\|F\|_\infty^2)\Big]\,4\,\mathbb{E}\Big[\exp\big((t_n - t_{n-1})K_T^2 W(T)^2\big)\Big] \\
&\qquad < \infty.
\end{aligned}$$

By Karatzas and Shreve (1991, Cor. 3.5.14) this implies the second property, whenever we choose $0 = t_0 < \cdots < t_N = T$ yielding finite expectations as above. □

Remarks 2.

- *As an alternative proof strategy the Girsanov theorem in Putschke (2001, Lemma 3.1.1) may be used, if the process $X(0) + W$ is understood as an affine stochastic delay process having the same initial function F, but zero weight measure and if the Radon-Nikodym derivative is restricted to the space $C([0,T])$.*
- *Every continuous Gaussian process F on $[-r,0]$ satisfies the moment condition due to the Fernique Theorem for Gaussian measures on $C([-r,0])$, see Da Prato and Zabczyk (1992, Thm. 2.6).*

Next, we consider two solution processes $X^{(1)}$ and $X^{(2)}$ commonly defined on our filtered probability space, keeping the measure $\mathbb{P}$ fixed, and derive their likelihood function. The main difficulty is due to the initial condition.

Corollary 2. *Let $X^{(1)}$ and $X^{(2)}$ be solutions of the affine SDDE (2.2.10) with weights a_1, a_2 and continuous initial functions F_1, F_2, respectively, where the moments $\mathbb{E}[\exp(\varepsilon\|F_1\|_\infty^2)]$ and $\mathbb{E}[\exp(\varepsilon\|F_2\|_\infty^2)]$ are both finite for some $\varepsilon > 0$. If the laws of $F_1(0)$ and $F_2(0)$ are mutually absolutely continuous, then the laws μ_1 and μ_2 on $C([0,T])$, induced by $X^{(1)}$ and $X^{(2)}$, are mutually absolutely continuous.*

With the convention of Remark 1 for functions $x \in C([0,T])$ the functionals

$$Z_t^{(1)}(x) = \mathbb{E}\left[\int_{-r}^0 X^{(1)}(t+s)da_1(s)\,\middle|\, \mathcal{F}_T^{X^{(1)}}\right](x),$$
$$Z_t^{(2)}(x) = \mathbb{E}\left[\int_{-r}^0 X^{(2)}(t+s)da_2(s)\,\middle|\, \mathcal{F}_T^{X^{(2)}}\right](x)$$

can be chosen such that $Z_t^{(i)}$ is $\mathcal{F}_t$-adapted and $Z_{\bullet_1}^{(i)}(\bullet_2)$ is jointly measurable with respect to the Borel-σ-algebra of the space $[0,T]\times C([0,T])$ for $i = 1,2$.

The Radon-Nikodym-derivative is then $\mathbb{P}$-a.s. given by

$$\Lambda_T(X^{(1)}, X^{(2)}) = \frac{d\mu_1}{d\mu_2} = \Lambda(F_1(0), F_2(0))\bullet$$
$$\bullet \exp\left(\int_0^T (Z_t^{(1)}(X^{(2)}) - Z_t^{(2)}(X^{(2)}))\,dW(t) - \frac{1}{2}\int_0^T (Z_t^{(1)}(X^{(2)}) - Z_t^{(2)}(X^{(2)}))^2 dt\right)$$

with $\Lambda(F_1(0), F_2(0)) = \frac{d\mathcal{L}(F_1(0))}{d\mathcal{L}(F_2(0))}$.

Proof. Since $F_1(0)$ and $F_2(0)$ have mutually absolutely continuous laws, the processes $F_1(0)+W$ and $F_2(0)+W$ induce equivalent laws on $C([0,T])$ with Radon-Nikodym derivative $\Lambda(F_1(0),F_2(0))$. This follows from the fact that for all bounded Borel-measurable functions $G:\mathbb{R}^{n+1}\to\mathbb{R}$ and all $0\le t_1\le\ldots\le t_n\le T$, $n\in\mathbb{N}$, by independence of F and W (put $\tilde{G}(x_0,x_1,\ldots,x_n):=G(x_0,x_1-x_0,\ldots,x_n-x_0)$)

$$\begin{aligned}
&\mathbb{E}[G(F_1(0),F_1(0)+W(t_1),\ldots,F_1(0)+W(t_n))]\\
&=\mathbb{E}[\tilde{G}(F_1(0),W(t_1),\ldots,W(t_n))]\\
&=\mathbb{E}[\tilde{G}(F_2(0),W(t_1),\ldots,W(t_n))\Lambda(F_1(0),F_2(0))]\\
&=\mathbb{E}[G(F_2(0),F_2(0)+W(t_1),\ldots,F_2(0)+W(t_n))\Lambda(F_1(0),F_2(0))]
\end{aligned}$$

holds and that the Borel-σ-algebra of $C([0,T])$ is generated by the coordinate projections. This kind of argument also provides the aforementioned adaptation to nonzero initial conditions. By Theorem 2 we thus conclude that μ_1 and μ_2 are mutually absolutely continuous.

From Liptser and Shiryaev (2001, Lemma 4.9) follows the asserted existence of the functionals $Z^{(1)}$ and $Z^{(2)}$. Moreover, from Thm. 7.13 in Liptser and Shiryaev (2001), adapted to nonzero initial conditions, follows $\mathbb{P}$-a.s. for $X=X^{(i)}$ and $Z=Z^{(i)}$, $i=1,2$, that X and the diffusion-type process Ξ, solving

$$d\Xi(t)=Z_t(\Xi)\,dt+dW(t),\quad t\ge 0,\qquad \Xi(0)=X(0),$$

induce the same law on $C([0,T])$, if the conditions of Theorem 2 and additionally

$$\int_0^T\mathbb{E}\left[\left|\int_{-r}^0 X(t+s)\,da(s)\right|\right]dt<\infty$$

are satisfied. For $X^{(1)}$ and $X^{(2)}$ this last condition is an immediate consequence of (2.2.11). Hence, the Radon-Nikodym derivatives agree Liptser and Shiryaev (2001, Eq. (7.83)):

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(-\int_0^T\int_{-r}^0 X(t+s)\,da(s)\,dX(t)+\frac{1}{2}\int_0^T\left(\int_{-r}^0 X(t+s)\,da(s)\right)^2dt\right)\,\middle|\,\mathcal{F}_T^X\right]\\
&=\exp\left(\int_0^T Z_t(X)\,dX(t)-\frac{1}{2}\int_0^T(Z_t(X))^2\,dt\right).
\end{aligned}$$

We find the Radon-Nikodym derivative $\frac{d\mu_1}{d\mu_2}$ under the law of $X^{(2)}$:

$$\begin{aligned}
&\Lambda_T(X^{(1)},X^{(2)})\\
&=\Lambda_T(X^{(1)},F_1(0)+W)\Lambda_T(F_1(0)+W,F_2(0)+W)\Lambda_T(F_2(0)+W,X^{(2)})\\
&=\Lambda(F_1(0),F_2(0))\exp\left(\int_0^T Z_t^{(1)}(X^{(2)})\,dX^{(2)}(t)-\frac{1}{2}\int_0^T(Z_t^{(1)}(X^{(2)}))^2\,dt\right)\bullet\\
&\quad\bullet\exp\left(-\int_0^T Z_t^{(2)}(X^{(2)})\,dX^{(2)}(t)+\frac{1}{2}\int_0^T(Z_t^{(1)}(X^{(2)}))^2\,dt\right)\\
&=\Lambda(F_1(0),F_2(0))\bullet\\
&\quad\bullet\exp\left(\int_0^T(Z_t^{(1)}-Z_t^{(2)})(X^{(2)})\,dW(t)-\frac{1}{2}\int_0^T((Z_t^{(1)}-Z_t^{(2)})(X^{(2)}))^2\,dt\right).
\end{aligned}$$

□

Remark 2. *The likelihood ratio $\Lambda_T(X^{(1)}, X^{(2)})$ for solutions with the same deterministic initial function $F_1 = F_2$ is given by*

$$\exp\left(\int_0^T \int_{-r}^0 X^{(2)}(t+s)\,da_{12}(s)\,dW(t) - \frac{1}{2}\int_0^T \left(\int_{-r}^0 X^{(2)}(t+s)\,da_{12}(s)\right)^2 dt\right)$$

with $a_{12} = a_1 - a_2$. Note that already for different deterministic functions F_1 and F_2 with $F_1(0) = F_2(0)$ the likelihood ratio is less simple. For the general case, the Markov property for the corresponding stochastic evolution equation on $C([-r, 0])$ Scheutzow (1983, Thm 2.1) yields that the conditional expectation in the definition of $Z^{(i)}$ for $T \geq r$ needs only to be taken with respect to $\mathcal{F}_r^{X^{(i)}}$, $i = 1, 2$, and we obtain

$$\Lambda_T(X^{(1)}, X^{(2)}) = \Lambda_r(X^{(1)}, X^{(2)}) \bullet$$
$$\exp\left(\int_r^T \int_{-r}^0 X^{(2)}(t+s)\,da_{12}(s)\,dW(t) - \frac{1}{2}\int_r^T \left(\int_{-r}^0 X^{(2)}(t+s)\,da_{12}(s)\right)^2 dt\right).$$

2.3 Stationary solutions

Recall that a stochastic process $(X(t),\, t \geq 0)$ is stationary if for all $\tau_1, \ldots, \tau_n \geq 0$, $n \in \mathbb{N}$ and $s > 0$ the vectors $(X(\tau_1), \ldots, X(\tau_n))$ and $(X(\tau_1 + s), \ldots, X(\tau_n + s))$ have the same distribution.

Theorem 1. *The affine SDDE (2.2.10) with weight $a \in M([-r, 0])$ admits a stationary solution X if and only if a is an element of $M^-([-r, 0])$. This stationary solution is unique.*

Proof. This follows from the general result by Gushchin and Küchler (2000). □

More abstractly, the theory of stochastic evolution equation Da Prato and Zabczyk (1992, Thm. 11.11) yields that for $a \in M^-$ there exists a unique invariant measure on the Delfour-Mitter Hilbert space $\mathbb{R} \times L^2([-r, 0])$, when the delay equation is lifted to this space, since then the transition operator T_t of the deterministic semigroup satisfies $\lim_{t\to\infty} \|T_t\| = 0$. The existence of a stationary law of the real valued process follows from the projection of this invariant measure onto its first coordinate. Though this approach does not immediately supply a proof of uniqueness, it is more general in the sense that the whole $L^2([-r, 0])$-valued segment process $(X_t)_{t\geq 0} = (X(t+s),\, -r \leq s \leq 0)_{t\geq 0}$ is stationary.

We gather some facts from Küchler and Mensch (1992) and Mohammed and Scheutzow (1990). A nice representation of the stationary solution process can be obtained by extending the Brownian motion to the whole real line and by setting

$$X(t) = \int_{-\infty}^t x_0(t-s)\,dW(s), \quad t \geq 0. \tag{2.3.13}$$

Thus, X is a moving average over the fundamental solution, which is in close analogy to the representation of stationary autoregressive processes Brockwell and Davis (1996). Note that the integral is well-defined due to $x_0 \in L^2([0, \infty))$. X is uniquely determined by the fact that it is a stationary Gaussian process which is centred and has (auto-)covariance function

$$q_a(t) := q(t) := \mathbb{E}[X(0)X(|t|)] = \int_0^\infty x_0(s)x_0(s+|t|)\,ds, \quad t \in \mathbb{R}. \tag{2.3.14}$$

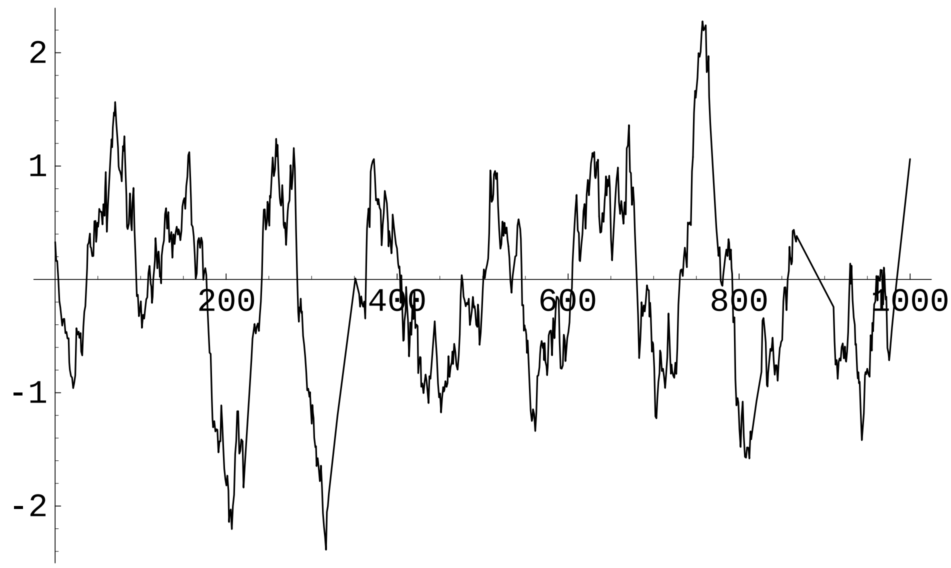

Figure 2.3.1: Trajectory of a stationary solution

From the properties of x_0, it follows that the covariance function q_a satisfies the linear deterministic delay equation

$$q_a'(t) = \int_{-r}^{0} q_a(t+s)\,da(s), \quad t \geq 0; \qquad q_a(t) = q_a(-t), \quad t \in [-r, 0]. \tag{2.3.15}$$

It is, by the way, quite remarkable that such symmetric solutions on $[-r, r]$ exist for linear delay equations. From the relationship (2.1.9) between the fundamental solution and the characteristic function a very useful explicit representation of the spectral density is obtained:

$$\hat{q}_a(\xi) := \int_{-\infty}^{\infty} q_a(t) e^{-i\xi t} dt = \frac{1}{|\chi_a(i\xi)|^2}, \quad \xi \in \mathbb{R}\,. \tag{2.3.16}$$

Note that due to $v_0(a) < 0$ the spectral density is continuous and decays like $|\xi|^{-2}$ for $|\xi| \to \infty$.

Example 2. *Consider again $a = \alpha\delta_0$. The solution to the affine SDDE is an Ornstein-Uhlenbeck process, which for $\alpha < 0$ admits a stationary solution with covariance function $q_{OU}(t) = \frac{1}{2|\alpha|} e^{\alpha t}$ and spectral density $\hat{q}_{OU}(\xi) = (\xi^2 + \alpha^2)^{-1}$. For the case of exponential weight functions see Proposition 2. Figure 2.3.1 shows a trajectory of the stationary solution process of the SDDE with weight function $g(t) = -10^{1+3t}$ for $t \in [-1, 0]$. By the Girsanov Theorem 2 this could also just be a path of Brownian motion.*

For the asymptotic study of the quality of the estimators we shall always assume the stationarity of the solution. That this assumption is justified in the stable regime $a \in M^-$, follows from the representation (2.2.12) which implies that any solution $X(t)$ of (2.2.10) converges exponentially fast in $L^2(\Omega)$ for $t \to \infty$ to the stationary solution at time t Mohammed et al. (1986, Thm. 3). The fact that the law of $X(t)$ converges to the stationary distribution for $t \to \infty$ is already a consequence of Da Prato and Zabczyk (1992, Thm. 11.11).

Another manifestation of this asymptotic independence of $X(t)$ from the initial process is the following mixing property of the stationary solution.

Theorem 3. *The stationary solution X of (2.2.10) with an M^--weight a is β-mixing; more precisely, for any $\delta \in (0, -v_0(a))$ and $\tau > 0$ there exists a $C > 0$ with*

$$\beta(t) := \mathbb{E}_a\left[\sup_{A\in\sigma(X(s),\,s\geq\tau+t)} |\mathbb{P}_a(A\,|\,\sigma(X(s),\,s\leq\tau)) - \mathbb{P}_a(A)|\right] \leq Ce^{-\delta t}$$

for all $t > 0$. For fixed $R > 0$ and $\varepsilon > 0$ the constant C may be chosen uniformly for all weights a from $M(R, \delta+\varepsilon)$.

Remarks 3.

- *Following Ibragimov and Rozanov (1978, Chap. IV), the β-mixing property is called absolute regularity and is for Gaussian processes equivalent to information regularity. Moreover, it implies the strong mixing or α-mixing property and, in our Gaussian setting, also the complete regularity*

$$r(t) := \sup_{\eta_1,\eta_2} \mathrm{Cov}_a[\eta_1, \eta_2] \leq 2\pi Ce^{-\delta t}, \quad t > 0, \tag{2.3.17}$$

  where the supremum is taken over all $\sigma(X(s),\, s\leq\tau)$-measurable centred random variables η_1 and $\sigma(X(s),\, s\geq\tau+t)$-measurable centred random variables η_2 with $\mathbb{E}_a[\eta_1^2] = \mathbb{E}_a[\eta_2^2] = 1$ and $\tau \geq 0$ arbitrary.

- *A Gaussian process is uniformly strongly mixing or φ-mixing if and only if the covariance function has compact support Ibragimov and Rozanov (1978, Thm. IV.5). This cannot happen in our case, since there do not exist any so called small solutions in the one-dimensional case Diekmann et al. (1995, Thm. V.4.3). Therefore, the β-mixing property is the strongest regularity property among those discussed by Ibragimov and Rozanov (1978) that X can satisfy.*

- *Another surprising consequence of the theorem, which strengthens an analytical result in Diekmann et al. (1995, p. 219), follows from the equivalence in Ibragimov and Rozanov (1978, Thm. VI.6). The complete regularity of X implies*

$$\sup_{\lambda\in\mathbb{R}} \sum_{\lambda_j} \left|\mathrm{Re}\left(\frac{1}{i\lambda - \lambda_j}\right)\right| < \infty,$$

  where the sum is taken over all zeros λ_j of the characteristic function χ.

- *The mixing property of X is not completely obvious, since stationary Gaussian processes with exponentially decreasing covariance functions exist that are not mixing at all. An example is provided by the Gaussian process Z with covariance function $\mathbb{E}[Z(0)Z(t)] = \exp(-t^2/2)$ and hence spectral density of the same form which prevents Z even from being regular, i.e. ergodic Ibragimov and Rozanov (1978, p. 113).*

Proof. According to Ibragimov and Rozanov (1978, Thm. IV.9), taking into account the equivalence of β-mixing and information regularity (loc.cit., p. 128), the following three conditions on the spectral density $\hat{q}_a = |\chi_a(i\bullet)|^{-2}$ have to be checked:

1. $\bar{\chi}_a(i\bullet)^{-1} = \chi_a(-i\bullet)^{-1}$ is a function in the Hardy space $\mathcal{H}^2$;
2. $\int_{-\infty}^{\infty}(1+\xi^2)^{-1}\log(\hat{q}_a(\xi))\,d\xi > -\infty$;
3. $I(T) := \limsup_{\varepsilon\to 0}\int_{-\infty}^{-T}|t||c_\varepsilon(t)|^2\,dt \lesssim e^{-2\delta T}$, $T > 0$, where c_ε is defined to be the Fourier transform

$$c_\varepsilon(t) = \mathcal{F}\left(\frac{i}{i+\varepsilon\bullet}\frac{\chi_a(i\bullet)}{\bar{\chi}_a(i\bullet)}\right)(t) = \int_{-\infty}^{\infty}\frac{i}{i+\varepsilon\lambda}\frac{\chi_a(i\lambda)}{\bar{\chi}_a(i\lambda)}e^{-i\lambda t}\,d\lambda.$$

   The bound for I must hold uniformly for weights $a \in M(R, \delta)$.

The first property follows from the spectral representation (2.1.9) and a Paley-Wiener theorem Ibragimov and Rozanov (1978, p. 35) since $x_0 \in L^2([0,\infty))$ holds for $a \in M^-$. For the second condition observe that $\hat{q}_a$ is continuous and that $|\hat{q}_a(\xi)| \sim (1+\xi^2)^{-1}$ holds due to (2.3.16).

For the third condition it suffices to establish

$$\limsup_{\varepsilon\to 0} \int_{-\infty}^{\infty} |te^{-\delta t} c_\varepsilon(t)|^2 \, dt =: K < \infty$$

with a uniform constant K for the weights in $M(R,\delta)$, since then $I(T) \leq KT^{-1}e^{-2\delta T}$ for $T > 0$ follows from $|te^{-2\delta t}| \geq Te^{2\delta T}$ for $t \leq -T$. The Plancherel identity and general properties of the complex Fourier transform for functions holomorphic in the strip $\{z \in \mathbb{C} \mid |\mathrm{Re}(z)| < v_0\}$ Katznelson (1976, Thm. VI.1.5 and Section VI.7.1) yield

$$K = 2\pi \limsup_{\varepsilon\to 0} \int_{-\infty}^{\infty} \left| \left(\frac{i}{i+\varepsilon\bullet} \frac{\chi_a(i\bullet)}{\bar{\chi}_a(i\bullet)} \right)' (\lambda + i\delta) \right|^2 d\lambda.$$

The estimate $|(fg)'|^2 \leq 2(|f'g|^2 + |fg'|^2)$, which is derived from the product rule, allows separate estimates. The first summand vanishes:

$$\begin{aligned}
&\limsup_{\varepsilon\to 0} \int_{-\infty}^{\infty} \left| \left(\frac{i}{i+\varepsilon(\bullet+i\delta)} \right)' (\lambda) \frac{\chi_a(i(\lambda+i\delta))}{\bar{\chi}_a(i(\lambda+i\delta))} \right|^2 d\lambda \\
&\qquad = \limsup_{\varepsilon\to 0} \int_{-\infty}^{\infty} \frac{\varepsilon^2}{|i+\varepsilon\lambda+i\varepsilon\delta|^4} \, d\lambda \leq \limsup_{\varepsilon\to 0} \int_{-\infty}^{\infty} \frac{\varepsilon}{|i+\lambda|^4} \, d\lambda = 0.
\end{aligned}$$

For the second summand the decay properties of χ_a come into play:

$$\begin{aligned}
&\limsup_{\varepsilon\to 0} \int_{-\infty}^{\infty} \left| \frac{i}{i+\varepsilon(\lambda+i\delta)} \left(\frac{\chi_a(i(\bullet+i\delta))}{\bar{\chi}_a(i(\bullet+i\delta))} \right)' (\lambda) \right|^2 d\lambda \\
&\qquad \leq \int_{-\infty}^{\infty} \frac{|\chi_a'(i\lambda-\delta)\bar{\chi}_a(i\lambda-\delta) - \chi_a(i\lambda-\delta)\bar{\chi}_a'(i\lambda-\delta)|^2}{|\bar{\chi}_a(i\lambda-\delta)|^4} \, d\lambda \\
&\qquad \leq \int_{-\infty}^{\infty} \frac{4|\chi_a'(i\lambda-\delta)|^2}{|\chi_a(i\lambda-\delta)|^2} \, d\lambda \\
&\qquad = \int_{-\infty}^{\infty} \frac{4|1-\int_{-r}^0 se^{(i\lambda-\delta)s} da(s)|^2}{|\chi_a(i\lambda-\delta)|^2} \, d\lambda \\
&\qquad \leq \int_{-\infty}^{\infty} \frac{4(1+re^{\delta r}\|a\|_{TV})^2}{|\chi_a(i\lambda-\delta)|^2} \, d\lambda \\
&\qquad \lesssim \int_{-\infty}^{\infty} |\chi_a(i\lambda-\delta)|^{-2} \, d\lambda
\end{aligned}$$

with a constant depending on δ and $\|a\|_{TV}$. Due to (2.1.6), $|\chi_a(i\lambda-\delta)|^{-2} \sim (1+\lambda^2)^{-1}$ is satisfied for $\lambda \in \mathbb{R}$ such that $K < \infty$ holds as required. Estimate (2.1.6) further proves that K may be chosen uniformly for $a \in M(R,\delta+\varepsilon)$. For we have $|\chi_a(i\lambda-\delta)|^{-2} \leq \frac{1}{4}|\lambda|^{-2}$ for $|\lambda| \geq 2\|a\|_{TV}$ and the integral

$$\int_{-2R}^{2R} |\chi_a(i\lambda-\delta)|^{-2} \, d\lambda$$

depends weakly continuously on a by Lemma 2 and $M(R,\delta+\varepsilon)$ is weakly compact (Corollary 1). Hence, the third condition holds uniformly. $\square$

A key ingredient of the subsequent discussions will be the regularity and the parameter-dependence of the covariance function q_a. By a dominated convergence argument, using (2.1.7) and (2.1.8), one derives

$$q'(0+) = \lim_{h\downarrow 0} h^{-1}\int_0^\infty x_0(s)x_0(h+s)\,ds = \int_0^\infty x_0(s)x_0'(s)\,ds = -\tfrac{1}{2}x_0(0)^2 = -\tfrac{1}{2}, \tag{2.3.18}$$

so that q' has a jump of size -1 at zero by the symmetry of q. As will be shown now, the covariance function will be quite regular elsewhere. For the notion of Sobolev spaces H^α we refer to Appendix A.1.

Proposition 1. *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in M^- with covariance functions q_n according to (2.3.14) and suppose $a_n \xrightarrow{w} a$ with $v_0(a) < -\delta < 0$. For $f:\mathbb{R}\to\mathbb{R}$ set $E_\delta(f)(t) := f(t)e^{\delta t}$, $t\in\mathbb{R}$.*

Then $\|E_\delta(q_n - q_a)\|_{H^\alpha(\mathbb{R})} \to 0$ holds for all $\alpha < \frac{5}{2}$. For any $\mu,\nu \in M^-$ with $v_0(\mu), v_0(\nu) < -\delta < 0$ the difference $E_\delta(q_\mu - q_\nu)$ is an element of $H^\alpha(\mathbb{R})$ for all $\alpha < \frac{5}{2}$.

Proof. In view of (2.3.16) we consider first the associated characteristic functions χ_n and χ_a. Due to $|\chi(i\xi)|^2 = \chi(i\xi)\chi(-i\xi)$ the Fourier transform $\hat{q}$ can be extended to a holomorphic function in the strip $\{z\in\mathbb{C} \mid |\mathrm{Im}(z)| < v_0\}$ Katznelson (1976, Section VI.7.1) and satisfies

$$\mathcal{F}(E_\delta(q))(\xi) = \hat{q}(\xi + i\delta) = \chi(i\xi-\delta)^{-1}\chi(-i\xi+\delta)^{-1}.$$

Setting $S := e^{\delta r}\sup_n \|a_n\|_{TV} < \infty$, we obtain from (2.1.6) $|\chi_n(\pm i\xi \mp \delta)|^{-1} \le (|\xi|-S)^{-1}$, $|\xi|\to\infty$. Lemma 2 in combination with the classical result from calculus about the convergence of maxima on compact sets yields for the respective choice of signs

$$\lim_{n\to\infty}\max_{|\xi|\le 2S}|\chi_n(\pm i\xi\mp\delta)|^{-1} = \max_{|\xi|\le 2S}|\chi_a(\pm i\xi\mp\delta)|^{-1} < \infty.$$

Due to $v_0(a_n)\to v_0(a) < -\delta$ (Theorem 1) the functions $\chi_n^{-1}(i\bullet-\delta)$ and $\chi_n^{-1}(-i\bullet+\delta)$ have no poles on the real axis for $n\ge N$, N sufficiently large, and are thus uniformly dominated for $n \ge N$:

$$|\chi_n(i\xi-\delta)|^{-1} \le M(\xi) \text{ and } |\chi_n(-i\xi+\delta)|^{-1} \le M(\xi), \quad \xi\in\mathbb{R},$$

with the finite function

$$M(\xi) := \begin{cases} \sup_{n\ge N}\max_{|\xi|\le 2S}\big(|\chi_n(i\xi-\delta)|^{-1} + |\chi_n(-i\xi+\delta)|^{-1}\big), & |\xi|\le 2S,\\ (|\xi|-S)^{-1}, & \text{otherwise.}\end{cases}$$

The inverse triangle inequality provides for $n\ge N$ the pointwise estimate

$$\begin{aligned}
&|\hat{q}_a(\xi+i\delta) - \hat{q}_n(\xi+i\delta)|\\
&\le \frac{|\chi_a(i\xi-\delta)|(|\chi_a(-i\xi+\delta)-\chi_n(-i\xi+\delta)|)}{|\chi_a(i\xi-\delta)\chi_a(-i\xi+\delta)\chi_n(i\xi-\delta)\chi_n(-i\xi+\delta)|}\\
&\quad + \frac{|\chi_n(-i\xi+\delta)|(|\chi_a(i\xi-\delta)-\chi_n(i\xi-\delta)|)}{|\chi_a(i\xi-\delta)\chi_a(-i\xi+\delta)\chi_n(i\xi-\delta)\chi_n(-i\xi+\delta)|}\\
&\le (|i\xi+\delta|+S)M^4(\xi)\big(|\chi_a(-i\xi+\delta)-\chi_n(-i\xi+\delta)| + |\chi_a(i\xi-\delta)-\chi_n(i\xi-\delta)|\big)\\
&\lesssim M^3(\xi)\big(|\chi_a(-i\xi+\delta)-\chi_n(-i\xi+\delta)| + |\chi_a(i\xi-\delta)-\chi_n(i\xi-\delta)|\big),
\end{aligned} \tag{2.3.19}$$

with a constant depending on S and δ. Observe that $|\chi_a(i\bullet\pm\delta) - \chi_n(i\bullet\pm\delta)|$ is uniformly bounded by $2S$. By Lemma 2 on pointwise convergence and due to $(1+\xi^2)^\alpha M^6(\xi)\in L^1(\mathbb{R})$ for $\alpha < \frac{5}{2}$ the dominated convergence theorem yields

$$\lim_{n\to\infty}\|(1+\xi^2)^\alpha|\mathcal{F}(E_\delta(q_a-q_n))|^2\|_{L^1(\mathbb{R})} = 0.$$

From the characterisation of Sobolev spaces by the Fourier transform (Appendix A.1) we conclude $E_\delta(q_a - q_n) \to 0$ in $H^\alpha(\mathbb{R})$. Even without any convergence assumption, the deduction (2.3.19) shows that $E_\delta(q_\mu - q_\nu)$ is bounded in $H^\alpha(\mathbb{R})$-norm for $v_0(\mu), v_0(\nu) < -\delta$. □

Corollary 3. *The restriction on $\mathbb{R}^+$ of any covariance function q in (2.3.14) satisfies $q|_{\mathbb{R}^+} \in H^\alpha(\mathbb{R}^+)$ for all $\alpha < \frac{5}{2}$, but $q \in H^\alpha(\mathbb{R})$ holds only for $\alpha < \frac{3}{2}$. The covariance function is Lipschitz continuous, i.e. $q \in C^{0,1}(\mathbb{R})$, and even $q|_{\mathbb{R}^+} \in C^{1,1}(\mathbb{R}^+)$ holds. For $\delta < -v_0$ the following estimates are valid:*

$$\sup_{t\geq 0}|q(t+h) - q(t)|e^{\delta t} \leq C_1 h, \quad \sup_{t>0}|q'(t+h) - q'(t)|e^{\delta t} \leq C_2 h, \quad h > 0.$$

For fixed $R > 0$, $\varepsilon > 0$ the constants C_1 and C_2 may be chosen to hold uniformly for weights a from $M(R, \delta + \varepsilon)$ and to be independent of h.

Proof. Observe that $q_{OU}(t) = \frac{1}{2}e^{-|t|}$ (cf. Example 2) lies in $H^\alpha(\mathbb{R})$ if and only if $\alpha < \frac{3}{2}$ holds. Therefore any covariance function q lies only in $H^\alpha(\mathbb{R})$ for $\alpha < \frac{3}{2}$ since q_{OU} has this property and $q - q_{OU} \in H^{3/2}(\mathbb{R})$ holds by Proposition 1. On $\mathbb{R}^+$, however, q_{OU} is arbitrarily often differentiable so that $q|_{\mathbb{R}^+} \in H^\alpha(\mathbb{R}^+)$, $\alpha < \frac{5}{2}$, follows immediately from $q - q_{OU} \in H^\alpha(\mathbb{R})$.

The functions q_{OU} and $q - q_{OU}$ are both Lipschitz continuous functions on $\mathbb{R}$ owing to $H^2(\mathbb{R}) \subset C^{0,1}(\mathbb{R})$ (Appendix A.1), hence q is Lipschitz continuous. In addition, $a_n \xrightarrow{w} a$ with $v_0(a) < -\delta$ implies by Proposition 1 and by Sobolev embeddings $\|E_\delta(q_{a_n} - q_a)\|_{C^{0,1}(\mathbb{R})} \to 0$ so that the Lipschitz constant of $E_\delta(q)$ depends weakly continuously on the weight. By Corollary 1 this constant can be chosen uniformly over $M(R, \delta+\varepsilon)$. The first estimate now follows from the triangle inequality and the Lipschitz continuity of $e^{\delta \bullet}$:

$$\begin{aligned} |q(t+h) - q(t)|e^{\delta t} &\leq |q(t+h)e^{\delta(t+h)} - q(t)e^{\delta t}| + |q(t+h)e^{\delta(t+h)}|(1 - e^{-\delta h}) \\ &\leq \|E_\delta(q)\|_{C^{0,1}}(h + 1 - e^{-\delta h}), \quad t,\, h > 0. \end{aligned}$$

To study the Lipschitz continuity of the derivative we use the fact that q_a satisfies the deterministic delay equation (2.3.15) and infer similarly

$$\begin{aligned} |q_a'(t+h) - q_a'(t)|e^{\delta t} &= \left|\int_{-r}^{0} (q_a(t+h+s) - q_a(t+s))\, da(s)\right| e^{\delta t} \\ &\leq 2\|a\|_{TV}\|q_a\|_{C^{0,1}([t-r,t+h])} h e^{\delta t} \lesssim h, \quad t,\, h > 0 \end{aligned}$$

with a constant only depending on $\|a\|_{TV}$ and $\|E_\delta(q_a)\|_{C^{0,1}}$, which can thus also be chosen uniformly. □

Another two corollaries show how these convergence properties of the covariance function are related with the actual processes.

Corollary 4. *Let $R > 0$ and $\delta > 0$ be fixed. Then the laws of the stationary solutions $\{X^{(a)} \,|\, a \in M(R, \delta)\}$ of the affine SDDE (2.2.10) with weight a are uniformly tight on $C([-r, \infty))$ with the topology of uniform convergence on compact sets.*

Proof. By Karatzas and Shreve (1991, p. 64) it suffices to check the two uniform moment conditions

$$\sup_{a\in M(R,\delta)} \mathbb{E}[X^{(a)}(-r)^2] < \infty,$$

$$\sup_{a\in M(R,\delta)} \sup_{t,s\in[0,T]} \mathbb{E}[(X^{(a)}(t) - X^{(a)}(s))^4] \leq C_T|t-s|^2, \quad C_T < \infty \text{ for all } T > 0.$$

Since $X^{(a)}$ is Gaussian, we simply need

$$\sup_{a\in M(R,\delta)} q_a(0) < \infty,$$
$$\sup_{a\in M(R,\delta)} \sup_{t,s\in[0,T]} 6(q_a(0) - q_a(t-s))^2 \le C_T|t-s|^2, \quad C_T < \infty \text{ for all } T > 0.$$

By the weak compactness of $M(R,\delta)$ (Corollary 1) the first condition follows from Proposition 1 and the second from the uniform Lipschitz continuity of q_a derived in Corollary 3. □

Corollary 5. *Suppose the sequence $(a_n) \subset M^-$ converges weakly to the weight $a_0 \in M^-$. Then for any $p \ge 1$ the moments of the likelihood ratio of the associated processes on $C([0,r])$ converge to one:*

$$\lim_{n\to\infty} \mathbb{E}_{a_0}[\Lambda_r(X^{(a_n)}, X^{(a_0)})^p] = 1.$$

This is to be understood in the sense that for sufficiently large n the moment is finite and converges.

Proof. Due to Lemma 2 and the bounds $|\chi_{a_n}(i\xi)| \in [|\xi| - \|a_n\|_{TV}, |\xi| + \|a_n\|_{TV}]$, $\xi \in \mathbb{R}$, there are constants $c, C > 0$ such that for all $n \in \mathbb{N}$ and also for $n = 0$

$$c(1+\xi^2)^{-1} \le |\chi_{a_n}(i\xi)|^{-2} \le C(1+\xi^2)^{-1}, \quad \text{i.e. } \hat{q}_{a_n}(\xi) \sim (1+\xi^2)^{-1}, \quad \xi \in \mathbb{R},$$

holds. By Ibragimov and Rozanov (1978, Thm. III.13) this estimate for the spectral density implies that the Gaussian processes $X^{(a_n)}$ and $X^{(a_0)}$ have mutually absolutely continuous laws on $C([0,r])$ if and only if

$$\int_0^r \int_0^r (q_{a_n} - q_{a_0})''(t-s)^2 \, ds\, dt < \infty$$

is satisfied. Furthermore, by reconsidering the steps leading from the expression of the log-likelihood ratio in Thm. III.8 to Thm. III.13 of Ibragimov and Rozanov (1978), one obtains

$$\mathrm{Var}_{a_0}[\log(\Lambda_r(X^{(a_n)}, X^{(a_0)})] \lesssim \|q_{a_n} - q_{a_0}\|^2_{H^2([-r,r])}$$

with a constant independent of n. From Proposition 1 we thus infer

$$\lim_{n\to\infty} \mathrm{Var}_{a_0}[\log(\Lambda_r(X^{(a_n)}, X^{(a_0)}))] = 0.$$

In terms of the operator $R^{(n)} := \mathrm{Id} - Q_{a_0}^{1/2} Q_{a_n}^{-1} Q_{a_0}^{1/2}$ with eigenvalues $(\lambda_i^{(n)})_{i\in\mathbb{N}}$ we obtain by using the eigenfunctions of $R^{(n)}$ (cf. Ibragimov and Rozanov (1978, Sec. III.2.1)) with independent real random variables $(\eta_i) \sim N(0,1)$:

$$\mathbb{E}_{a_0}[\Lambda_r(X^{(a_n)}, X^{(a_0)})^p] = \mathbb{E}\left[\prod_{i=1}^{\infty} (1-\lambda_i^{(n)})^{p/2} \exp(\tfrac{p}{2}\lambda_i^{(n)}\eta_i^2)\right].$$

The convergence of $\mathrm{Var}_{a_0}[\log(\Lambda_r(X^{(a_n)}, X^{(a_0)}))]$ is thus equivalent with the convergence of the Hilbert-Schmidt norm of the operators $R^{(n)}$, i.e.

$$\lim_{n\to\infty} \sum_{i=1}^{\infty} (\lambda_i^{(n)})^2 = 0$$

holds. By the independence of (η_i) we conclude

$$\begin{aligned}
\log(\mathbb{E}_{a_0}[\Lambda_r(X^{(a_n)},X^{(a_0)})^p]) &= \sum_{i=1}^{\infty}\Big(\tfrac{p}{2}\log(1-\lambda_i^{(n)})+\log((1-p\lambda_i^{(n)})^{-1/2})\Big)\\
&= \frac{1}{2}\sum_{i=1}^{\infty}\Big(p\log(1-\lambda_i^{(n)})-\log((1-p\lambda_i^{(n)})\Big)\\
&\sim \frac{1}{2}\sum_{i=1}^{\infty}\Big(p(-\lambda_i^{(n)}-\tfrac{1}{2}(\lambda_i^{(n)})^2)-(-p\lambda_i^{(n)}-\tfrac{1}{2}(p\lambda_i^{(n)})^2)\Big)\\
&= \sum_{i=1}^{\infty}\tfrac{p(p-1)}{2}(\lambda_i^{(n)})^2\\
&\to 0 \text{ for } n\to\infty.
\end{aligned}$$

In the first line we used the identity $\mathbb{E}[\exp(\alpha\eta_i^2)]=(1-2\alpha)^{-1/2}$ for $\alpha\in[0,\frac{1}{2})$ and in the third line the Taylor expansion of $\log(1+\bullet)$. In fact, the estimate $\log(1-x)+x+\frac{1}{2}x^2\in[-\varepsilon x^2,\varepsilon x^2]$ for all $\varepsilon>0$ and $|x|\le K(\varepsilon)$ together with $\lim_n\sup_i|\lambda_i^{(n)}|\to 0$ had been used for the precise argument in the third line. □

2.4 Case study: exponential weight function

We focus on weight measures with exponential density of the form

$$da(s)=-\beta e^{\alpha s}\,ds,\quad \alpha,\,\beta\in\mathbb{R},$$

modeling a memory effect that is waxing ($\alpha<0$) or waning ($\alpha>0$) with a constant rate. The case $\alpha=0$ yields an arithmetic mean over the past. From a mathematical perspective an exponential weight function is very attractive because it often allows to transform infinite-dimensional problems to finite-dimensional ones (c.f. Elsanosi et al. (2000) for bounded memory, Scheutzow (1983) for unbounded memory).

There is a large literature on stability regions, i.e. parameter values yielding $v_0<0$, for the weights with point delay (Bellman and Cooke (1963), Diekmann et al. (1995)). The covariance function in the case of a one point delay has been investigated by Küchler and Mensch (1992). We shall first investigate the region of stability in the parameter plane for the exponential weight functions with bounded delay $0<r<\infty$ and then determine the corresponding covariance functions.

The stability region

We have to study the zeros of the characteristic function

$$\chi_{\alpha,\beta,r}(\lambda):=\lambda-\int_{-r}^{0}e^{\lambda s}(-\beta)e^{\alpha s}\,ds\overset{(\alpha+\lambda\neq 0)}{=}\lambda+\beta\frac{1-e^{-(\alpha+\lambda)r}}{\alpha+\lambda}.$$

It should be mentioned that a similar analysis of the zeros of a function of this type has been carried out in Bellman and Cooke (1963, Thm. 13.9). Nevertheless we would like to cover also the case $\alpha\le 0$ and strive for asymptotic results as well, whereas Bellman and Cooke (1963) only state the result in an abstract manner not yielding numerical values.

From Lemma 1 we infer $v_0\ge 0$ for $\beta\le 0$ and we therefore assume $\beta>0$ in the sequel. The substitutions $\tilde\alpha:=\alpha r$, $\tilde\beta:=\beta r^2$ and $\tilde\lambda:=\lambda r$ yield

$$\chi_{\alpha,\beta,r}(\lambda)\overset{(\alpha+\lambda\neq 0)}{=}\left(\tilde\lambda+\tilde\beta\frac{1-e^{-(\tilde\alpha+\tilde\lambda)}}{\tilde\alpha+\tilde\lambda}\right)r^{-1},$$

whence without loss of generality we can assume $r = 1$ for the search of the stability region. We henceforth consider

$$da(s) = -\beta e^{\alpha s}\, ds \quad \text{and} \quad \chi_{\alpha,\beta}(\lambda) = \begin{cases} \frac{\lambda^2+\alpha\lambda+\beta-\beta e^{-\alpha}e^{-\lambda}}{\alpha+\lambda}, & \alpha+\lambda \neq 0 \\ \lambda+\beta, & \alpha+\lambda = 0 \end{cases} \tag{2.4.20}$$

for $r = 1$, $\alpha \in \mathbb{R}$ and $\beta > 0$.

In order to determine the number of zeros of $\chi_{\alpha,\beta}$ in the right half plane, we need a result from complex analysis.

Lemma 3. *Let $F : \mathbb{R}^2 \times \mathbb{C} \to \mathbb{C}$ be a continuous function such that for all $(\alpha, \beta) \in \mathbb{R}^2$ the function $z \mapsto F(\alpha, \beta, z)$ is holomorphic on $\mathbb{C}$. Let $O \subset \mathbb{C}$ be an open and bounded set and let (α_0, β_0) be such that $F(\alpha_0, \beta_0, z) \neq 0$ holds for all z on the boundary ∂O of O. Then there exists a neighbourhood $U \subset \mathbb{R}^2$ of (α_0, β_0) such that*

- *$F(\alpha, \beta, z) \neq 0$ holds for all $(\alpha, \beta) \in U$ and all $z \in \partial O$;*
- *the number of zeros of $F(\alpha, \beta, \bullet)$ in O, taking multiplicities into account, is constant.*

In other words, the set

$$N_n(O) := \{(\alpha, \beta) \in \mathbb{R}^2 \mid F(\alpha, \beta, \bullet) \neq 0 \text{ on } \partial O,\ F(\alpha, \beta, \bullet) \text{ has } n \text{ zeros in } O\}$$

is an open subset of $\mathbb{R}^2$ for all $n \in \mathbb{N}_0$.

Proof. This is Diekmann et al. (1995, Lemma 2.8), which relies on the argument principle in complex analysis. □

We shall apply this lemma with $F(\alpha, \beta, z) = \chi_{\alpha,\beta}(z)$, which obviously satisfies all the conditions imposed. The right half plane $\{z \in \mathbb{C} \mid \mathrm{Re}(z) > 0\}$ is not bounded, but due to (2.1.6) we only need to consider the zeros of $\chi_{\alpha,\beta}$ in some semicircle $O = \{z \in \mathbb{C} \mid \mathrm{Re}(z) > 0, |z| < R\}$ for R sufficiently large. Any $(\alpha, \beta) \in \mathbb{R}^2$ for which $\chi_{\alpha,\beta}$ does not vanish on the imaginary axis belongs to $N_n(O)$ for some $n \in \mathbb{N}_0$. The parameter plane $\mathbb{R}^2$ is thus partitioned into open sets, where the number of zeros of $\chi_{\alpha,\beta}$ is constant and no zero lies on ∂O, and into (closed) sets, where $\chi_{\alpha,\beta}$ has a zero on ∂O.

We shall determine the regions in the parameter plane where the number of zeros in O is constant. For this, we first treat the case $\alpha = 0$, then derive asymptotic results for the parameter values (α, β) corresponding to zeros on the imaginary axis and finally extend the case $\alpha = 0$ to the general case using the preceding lemma.

The case $\boldsymbol{\alpha = 0}$. The characteristic function simplifies to

$$\chi_{0,\beta}(\lambda) = \begin{cases} \frac{\lambda^2+\beta-\beta e^{-\lambda}}{\lambda}, & \lambda \neq 0 \\ \beta, & \lambda = 0 \end{cases}.$$

Due to $\beta > 0$ we need not consider $\lambda = 0$. Any purely imaginary zero $\lambda = iy$, $y \in \mathbb{R}\setminus\{0\}$, of $\chi_{0,\beta}$ satisfies the equations

$$\begin{aligned} -y^2 + \beta - \beta\cos(y) &= 0, \\ -\beta\sin(y) &= 0. \end{aligned}$$

Necessarily $y = k\pi$, $k \in \mathbb{Z}\setminus\{0\}$, and $\beta(1-(-1)^k) = k^2\pi^2$ hold, which imply further that k cannot be even. Hence

$$\{(y_k, \beta_k) \mid y_k = (2k+1)\pi,\ \beta_k = \tfrac{\pi^2}{2}(2k+1)^2, \quad k \in \mathbb{Z}\}$$

is the set of solutions of this system of equations. Lemma 3 therefore shows that the number of zeros of $\chi_{0,\beta}$ in O is constant as a function of β on each of the open intervals (β_k, β_{k+1}), $k \in \mathbb{N}_0$, and on $(0, \beta_0)$.

What happens at the boundary points of these intervals? Due to $\chi_{0,\beta_k}(\lambda_k) = 0$ this can be considered as an implicit function problem; we wish to know how the zero λ_k moves for small perturbations of β_k. Thus, think of the parameter $\beta = \beta(\lambda)$ to be dependent on the zero λ of $\chi_{0,\beta}$ and allow for complex values of β, then for $\lambda \in \mathbb{C} \setminus \{2k\pi i \mid k \in \mathbb{Z}\}$

$$\chi_{0,\beta(\lambda)}(\lambda) = \frac{\lambda^2 + \beta(\lambda) - \beta(\lambda)e^{-\lambda}}{\lambda} = 0 \iff \beta(\lambda) = \frac{\lambda^2}{e^{-\lambda} - 1}.$$

Thus, at a given point λ the characteristic function has a zero for the (complex-valued) parameter $\beta(\lambda)$. We obtain as derivative

$$\beta'(\lambda) = \frac{2\lambda(e^{-\lambda} - 1) + \lambda^2 e^{-\lambda}}{(e^{-\lambda} - 1)^2}.$$

At $\lambda_k = iy_k = (2k+1)\pi i$ this gives $\beta'(iy_k) = iy_k + \frac{1}{4}y_k^2 \neq 0$. By the existence of local inverses of holomorphic functions Ahlfors (1979, Sections 2.3, 3.3) we can define the inverse $\lambda(\beta)$ of $\beta(\lambda)$ locally around $\beta_k = \beta(iy_k)$ and its derivative at β_k is determined by

$$\lambda'(\beta_k) = \frac{1}{\beta'(\lambda_k)} = \frac{1}{iy_k + \frac{1}{4}y_k^2} = \frac{-iy_k + \frac{1}{4}y_k^2}{|iy_k + \frac{1}{4}y_k^2|^2}.$$

Hence, $\mathrm{Re}(\lambda'(\beta_k)) > 0$ holds and the zero $\lambda_k = \lambda(\beta_k)$ moves with increasing β into the right half plane. This shows that locally at β_k at least two complex conjugate zeros of $\chi_{0,\beta}$ cross the imaginary axis from the left to the right as β increases.

That there are not more than two zeros moving into the right half plane can be proved by considering $O_\delta := \{z \in \mathbb{C} \mid \mathrm{Re}(z) > -\delta, |z| < R\}$, $\delta > 0$ and $R = R(\delta)$ large enough (cf. (2.1.6)). We choose $\delta > 0$ so small that there is no zero of χ_{0,β_k} on the boundary ∂O_δ and such that χ_{0,β_k} on $O_\delta \setminus O$ vanishes only at λ_k and $\lambda_{-(k+1)}$. By Lemma 3 the number of zeros of $\chi_{0,\beta}$ in O_δ is locally constant at $\beta = \beta_k$ such that the number of zeros of $\chi_{0,\beta}$ in O can increase by at most two at the points β_k, $k \in \mathbb{N}_0$.

The first condition of Lemma 1 shows $v_0 < 0$ for $\beta \in (0, \frac{\pi}{2})$, which by Lemma 3 extends to the whole interval $(0, \beta_0)$. Altogether, we thus know that there exist exactly $2k$ zeros of $\chi_{0,\beta}$ with positive real part for $\beta \in (\beta_{k-1}, \beta_k)$, $k \in \mathbb{N}$. For a visualization we refer to the forthcoming Figures 2.4.2 and 2.4.3.

The case of arbitrary α. We also regard the parameter values for which the characteristic function has zeros on the imaginary axis. Splitting into real and imaginary part we obtain from (2.4.20) for a zero $\lambda = iy$, $y \in \mathbb{R}$, of $\chi_{\alpha,\beta}$:

$$y^2 = \beta(1 - e^{-\alpha}\cos(y)), \tag{2.4.21}$$

$$\alpha y = -\beta e^{-\alpha}\sin(y). \tag{2.4.22}$$

Note that $\alpha + iy = 0$ would imply $\alpha = y = 0$ and thus $\chi_{\alpha,\beta}(\lambda) = \beta \neq 0$. Furthermore, with $y \in \mathbb{R}$ also $-y$ is a solution and $y = 0$ would again imply $\alpha = 0$, whence we shall assume $y > 0$ in the sequel. An elimination of β yields

$$\alpha(e^\alpha - \cos(y)) = -y\sin(y). \tag{2.4.23}$$

As a function of y the left hand side remains bounded while the function on the right oscillates with an amplitude growing to infinity. Hence, there are for fixed

$\alpha \in \mathbb{R}$ infinitely many solutions $y_k(\alpha)$, $k \in \mathbb{N}$, ordered by magnitude $0 < y_1(\alpha) < y_2(\alpha) < \cdots$. We derive for $k \in \mathbb{N}$ the estimates

$$|\alpha|(e^\alpha - 1) \le y_k(\alpha)|\sin(y_k(\alpha))| \le |\alpha|(e^\alpha + 1), \tag{2.4.24}$$

$$|\alpha|(e^\alpha - 1) \le y_k(\alpha) < |\alpha|(e^\alpha + 1) + 2k\pi. \tag{2.4.25}$$

The upper bound for $y_k(\alpha)$ follows from the fact that by the 2π-periodicity of the sine function in the interval $[|\alpha|(e^\alpha+1)+2(k-1)\pi, |\alpha|(e^\alpha+1)+2k\pi)$, $k \in \mathbb{N}$, at least one solution y must exist. From (2.4.22) we infer for the values $\beta_k(\alpha)$ corresponding to $y_k(\alpha)$

$$\beta_k(\alpha) = -\frac{\alpha e^\alpha y_k(\alpha)}{\sin(y_k(\alpha))}, \quad k \in \mathbb{N}. \tag{2.4.26}$$

Hence for $\alpha > 0$

$$\frac{e^\alpha \alpha^2 (e^\alpha - 1)^2}{e^\alpha + 1} \le \beta_k(\alpha) = \frac{\alpha e^\alpha y_k(\alpha)^2}{y_k(\alpha)|\sin(y_k(\alpha))|} \le \frac{e^\alpha(\alpha(e^\alpha + 1) + 2k\pi)^2}{e^\alpha - 1} \tag{2.4.27}$$

holds, which shows $\beta_k(\alpha) \sim \alpha^2 e^{2\alpha}$ for $\alpha > 0$ and any fixed $k \in \mathbb{N}$. The estimates derived are not very useful for $\alpha < 0$. In this case we use that the left hand side of (2.4.23), as a function of y, changes its sign from minus to plus in the intervals $(2l\pi - \frac{\pi}{2}, 2l\pi)$, $l \in \mathbb{N}$, whereas the right hand side is positive at $2l\pi - \frac{\pi}{2}$ and zero at $2l\pi$, $l \in \mathbb{N}$. Hence, we conclude $y_k(\alpha) < 2k\pi$ for all $\alpha < 0$. From (2.4.23) then follows $e^\alpha - \cos(y_k(\alpha)) \to 0$ for $\alpha \to -\infty$ and fixed $k \in \mathbb{N}$, hence $|\sin(y_k(\alpha))| \to 1$. Equation (2.4.26) yields for $\alpha \to -\infty$ the asymptotics $\beta_k(\alpha) \sim |\alpha|e^\alpha$, $k \in \mathbb{N}$.

For $\alpha \to 0$ we can also determine the asymptotic behaviour of $\beta_k(\alpha)$. Equation (2.4.23) shows $y_k(\alpha)\sin(y_k(\alpha)) \to 0$, i.e. $y_k(\alpha) \to l\pi$ for some $l \in \mathbb{N}_0$, when $\alpha \to 0$; more precisely, we even have $y_k(\alpha) \to k\pi$, since for sufficiently small $|\alpha|$ the left hand side of (2.4.23) as well as its derivative with respect to y are arbitrarily close to zero, whence the multiplicity of the solution $y_k(\alpha)$ is one and there exists exactly one solution in a small neighbourhood of $l\pi$ for all $l \in \mathbb{N}$ (exclude $l = 0$: the left hand side of (2.4.23) is positive at $y = 0$ for all $\alpha \ne 0$, whereas the right hand side is not positive for $y \in [-\pi, \pi]$). From (2.4.21) we thus derive

$$\lim_{\alpha \to 0} \beta_k(\alpha) = \begin{cases} \frac{k^2\pi^2}{2}, & k \text{ odd}, \\ \infty, & k \text{ even}. \end{cases}$$

For small values of $\alpha > 0$ the lower bound in (2.4.27) is not sharp. It can be improved due to $y_k(\alpha) \ge \pi$ for $\alpha > 0$, which follows from a consideration of the signs in (2.4.23):

$$\beta_k(\alpha) = \frac{|\alpha| e^\alpha y_k(\alpha)^2}{y_k(\alpha)|\sin(y_k(\alpha))|} \ge \frac{e^\alpha \pi^2}{e^\alpha + 1} = \frac{\pi^2}{1 + e^{-\alpha}}, \quad \alpha > 0.$$

By extending the stability region from $\alpha = 0$ to general values of α by Lemma 3, we have determined several properties of the stability region.

Proposition 1. *The stability region with $v_0 < 0$ for (2.4.20) in the parameter plane $\mathbb{R}^2$ contains the region*

$$\{(\alpha, \beta) \,|\, \alpha \ge 0,\, \beta > 0,\, \beta(1 + e^{-\alpha}) < \max(\pi^2, \alpha^2(e^\alpha - 1)^2)\}.$$

Regarding the asymptotic behaviour, there are constants $C_1 > C_2 > 0$ such that the stability region S lies "sandwiched" between two sets:

$$S \subset \{(\alpha, \beta) \,|\, \alpha < 0,\, \beta \in (0, C_1|\alpha|e^\alpha)\} \cup \{(\alpha, \beta) \,|\, \alpha \ge 0,\, \beta \in (0, C_1\alpha^2 e^{2\alpha})\},$$

$$S \supset \{(\alpha, \beta) \,|\, \alpha < 0,\, \beta \in (0, C_2|\alpha|e^\alpha)\} \cup \{(\alpha, \beta) \,|\, \alpha \ge 0,\, \beta \in (0, C_2\alpha^2 e^{2\alpha})\}.$$

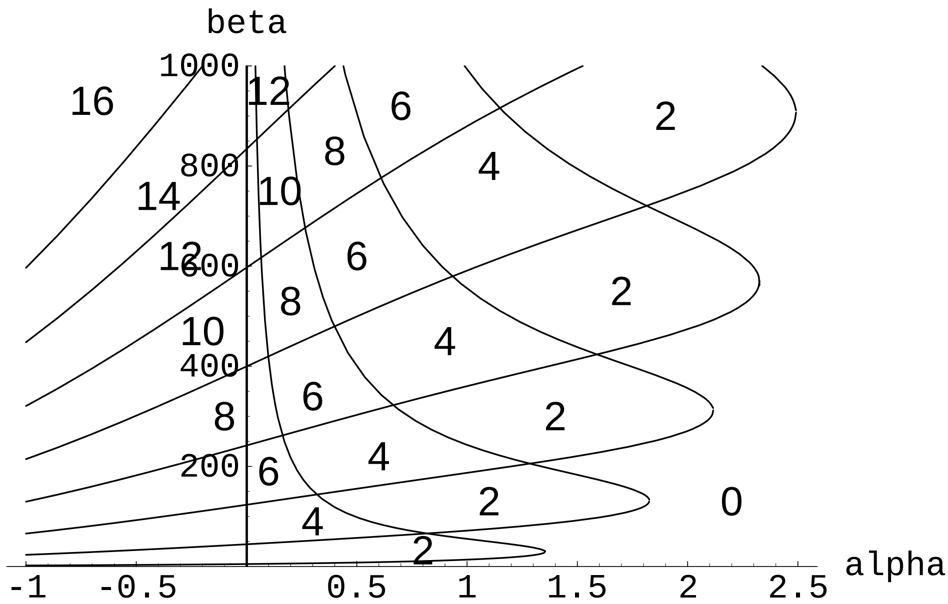

Figure 2.4.2: The number of zeros of $\chi_{\alpha,\beta}$ in (2.4.20) with positive real part

Remark 3. *It is interesting to compare the results with the case of unbounded delay ($r = \infty$). The analogously defined characteristic function changes to*

$$\chi(\lambda) = \lambda - \int_{-\infty}^{0} e^{\lambda s}(-\beta)e^{\alpha s}\,ds = \lambda + \frac{\beta}{\alpha+\lambda}, \quad \mathrm{Re}(\lambda) > -\alpha.$$

Exactly for $\alpha, \beta > 0$ its zeros have only negative real parts, whence it can be deduced that the stability region is the quadrant $\{(\alpha, \beta) \,|\, \alpha, \beta > 0\}$ Riedle (2001). Keeping $\alpha, \beta > 0$ fixed in our setting, we consider $r \to \infty$. Then from the reduction step at the beginning we have $\tilde{\alpha}_r = r\alpha$ and $\tilde{\beta}_r = r^2\beta$ and the parabola described by the curve $(\tilde{\alpha}_r, \tilde{\beta}_r)$ for $r \to \infty$ exponentially fast enters the stability region, because the sufficient condition $\tilde{\beta}_r < C_2\tilde{\alpha}_r^2 e^{2\tilde{\alpha}_r}$ is equivalent to $\beta < C_2\alpha^2 e^{2\alpha r}$.

A numerical computation of the curves in the parameter plane for which there are zeros of the characteristic function on the imaginary axis reveals a complicated nonlinear structure (Figures 2.4.2, 2.4.3). In particular, the stability region is not convex and not even isotonous in the sense that $v_0(a) < 0$ implies $v_0(\tau a) < 0$ for all $\tau \in (0, 1)$. This behaviour is rather counterintuitive. With regard to Figure 2.4.3 imagine for instance starting with $\alpha = 1$, $\beta = 100$. Then there are two zeros with positive real part. Since too strong negative feedback leads to exploding oscillations, one might reduce β to 60 and $v_0 < 0$ holds, but when the value of β is decreased further, the stability region is left again (e.g., for $\beta = 30$) before one returns to the stability region for small values of β. This kind of behaviour does not occur in the case of two point delays or any other of the case studies considered in Diekmann et al. (1995, Chap. XI).

The covariance function

We first reveal a very interesting feature. The solution x of the deterministic delay equation (2.1.1) with exponential weight function is also the solution of a linear

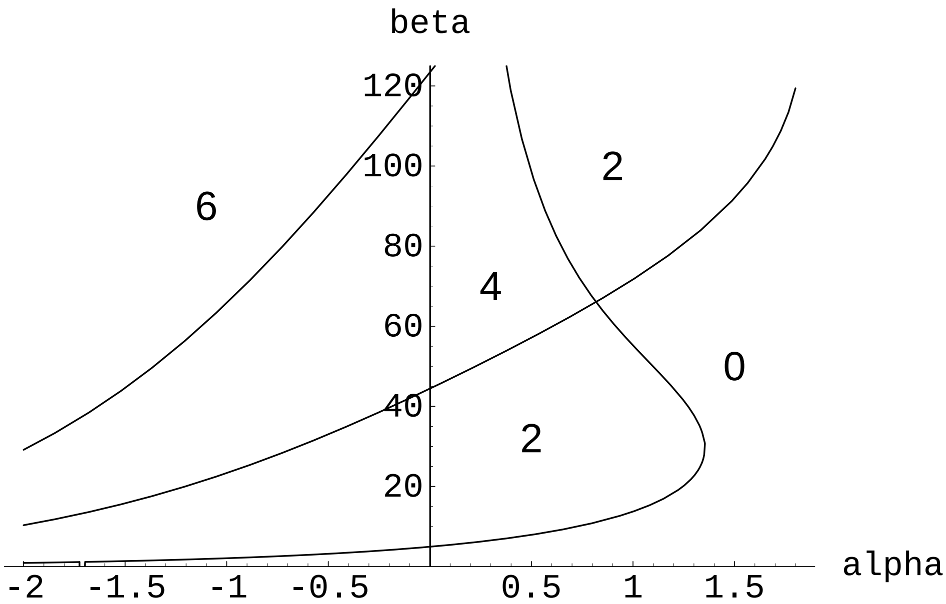

Figure 2.4.3: Detail of Figure 2.4.2

differential equation of second order with point delay:

$$x'(t) = -\beta \int_{-1}^{0} x(t+s)e^{\alpha s}\, ds = -\beta e^{-\alpha t} \int_{t-1}^{t} x(s)e^{\alpha s}\, ds, \quad t \geq 0,$$
$$x''(t) = -\beta(x(t) - x(t-1)e^{-\alpha}) - \alpha x'(t), \quad t \geq 0.$$

Also with regard to the problem of stability this opens an interesting new perspective. One might think of x describing an oscillating physical system with friction α and a back driving force of the form $\beta(x(t) - x(t-1)e^{-\alpha})$ exhibiting a memory effect. Already for modest values of $\alpha > 0$ the memory effect is rather small, but for $\alpha < 0$ there is not only a negative "friction" term, but also an in the mean forward driving force term, which together might result in a very unstable regime. It should be remarked that any weight function of the form

$$g(t) = \sum_{i=1}^{n} \gamma_i t^i e^{\alpha_i t}, \quad t \in [-r, 0], \quad n \in \mathbb{N},\, \gamma_1, \ldots, \gamma_n,\, \alpha_1, \ldots, \alpha_n \in \mathbb{R},$$

gives rise to a higher order differential equation only involving point delays at time $-r$, r being the memory length. This is proved analogously by multiple differentiation, a technique that is well known for unbounded memory ($r = \infty$), in which case the point delay disappears and an ordinary differential equation of higher order results (cf. Scheutzow (1983) and the references given there).

Using the fact that the covariance function q solves the deterministic delay equation (2.3.15), we obtain

$$q''(t) = -\beta(q(t) - e^{-\alpha} q(1-t)) - \alpha q'(t), \quad t \in [0, 1]. \tag{2.4.28}$$

We consider the complex-valued ansatz function

$$x(t) = A \sin(\omega(t - \tfrac{1}{2})) + B \cos(\omega(t - \tfrac{1}{2}))$$

with some $A, B, \omega \in \mathbb{C}$. Equation (2.4.28) and a comparison of the coefficients in front of the sine and cosine terms require for x to be a solution of (2.4.28) the conditions

$$\begin{aligned}-A\omega^2 &= -A\beta(1+e^{-\alpha}) + B\alpha\omega,\\ -B\omega^2 &= -B\beta(1-e^{-\alpha}) - A\alpha\omega.\end{aligned}$$

We aim at solving the linear system

$$\begin{pmatrix} -\omega^2+\beta(1+e^{-\alpha}) & \alpha\omega \\ \alpha\omega & -\omega^2+\beta(1-e^{-\alpha}) \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{2.4.29}$$

for non-trivial values of (A, B). This can be done if and only if the determinant of the matrix vanishes, hence

$$\omega^4 - (2\beta+\alpha^2)\omega^2 + \beta^2(1-e^{-2\alpha}) = 0,$$

or equivalently

$$\omega^2 = \beta + \tfrac{1}{2}\alpha^2 \pm \sqrt{(\beta+\tfrac{1}{2}\alpha^2)^2 - \beta^2(1-e^{-2\alpha})} \tag{2.4.30}$$

holds. The argument of the square root is always strictly positive due to $\beta > 0$, hence ω^2 is real and there are two different solutions for ω^2. We obtain exactly two solutions ω_+ and ω_- of (2.4.30) that are nonnegative ($\omega^2 \geq 0$) or have positive imaginary part ($\omega^2 < 0$). Note that $\alpha \geq 0$ implies $\omega^2 \geq 0$. Moreover, the matrix is never the zero matrix, because the upper left and lower right entry always differ. Hence, there are vectors $v_+ = (\xi_+, \eta_+)$, $v_- = (\xi_-, \eta_-)$ spanning the one-dimensional kernel of this matrix for $\omega = \omega_+$ and $\omega = \omega_-$, respectively, and we obtain two linearly independent solutions

$$\begin{aligned}x_+(t) &= \xi_+ \sin(\omega_+(t-\tfrac{1}{2})) + \eta_+ \cos(\omega_+(t-\tfrac{1}{2})),\\ x_-(t) &= \xi_- \sin(\omega_-(t-\tfrac{1}{2})) + \eta_- \cos(\omega_-(t-\tfrac{1}{2})).\end{aligned}$$

A general complex-valued solution of (2.4.28) is given by

$$x(t) = C_+x_+(t) + C_-x_-(t), \quad C_+, C_- \in \mathbb{C}\,. \tag{2.4.31}$$

From the theory of ordinary differential equations it is known that for the linear system

$$\begin{aligned}&f_1''(t) = -\beta(1-e^{-\alpha})f_1(t) - \alpha f_2'(t), \quad t \in (0,1),\\ &f_2''(t) = -\beta(1+e^{-\alpha})f_2(t) - \alpha f_1'(t), \quad t \in (0,1),\\ &f_2(\tfrac{1}{2}) = 0, \quad f_1'(\tfrac{1}{2}) = 0,\end{aligned} \tag{2.4.32}$$

the set of solution functions (f_1, f_2) forms a linear space of dimension two. On the other hand, any function y solving (2.4.28) gives rise to the solutions $f_1 := y + y(1-\bullet)$, $f_2 := y - y(1-\bullet)$ of (2.4.32) and the mapping $y \mapsto (f_1, f_2)$ is injective, which shows that the solution space of (2.4.28) is at most two dimensional and is given by the set of linear combinations (2.4.31). In particular, there are constants C_{q+} and C_{q-} such that $q = C_{q+}x_+ + C_{q-}x_-$ holds on $[0,1]$. How can these constants be found? We claim that by the two linear conditions on q (cf. Section 2.3)

$$q'(0+) = -\tfrac{1}{2} \text{ and } q'(t) = \int_{-r}^{0} q(|t+s|)(-\beta e^{\alpha s})\,ds \tag{2.4.33}$$

the constants C_{q+} and C_{q-} are uniquely determined. The easiest way to see this is to consider the functions

$$q_{ab}(t) = aq(t) + b\int_0^t q(t-s)\,ds, \quad t \in [0,1],\ a, b \in \mathbb{C},$$

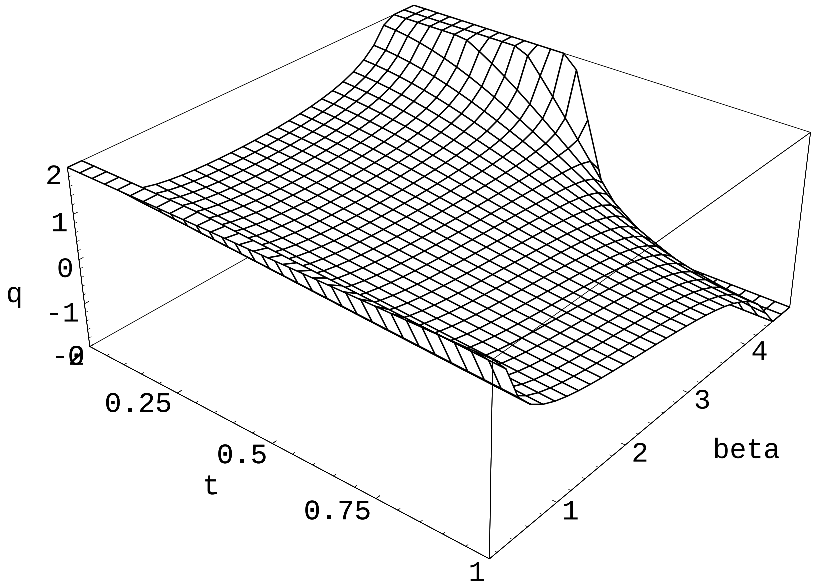

Figure 2.4.4: The covariance function q on $[0,1]$ for $\alpha = 0$ and $\beta \in (0, \frac{\pi^2}{2})$

which also satisfy (2.4.28) due to the variation of constants result

$$q'_{ab}(t) = \int_{-r}^{0} q_{ab}(|t+s|)(-\beta e^{\alpha s})\, ds + bq(0), \quad t \in [0,1].$$

Hence, they span the two-dimensional solution space of (2.4.28). Imposing additionally the second condition in (2.4.33) reduces the solution space to a one-dimensional space ($b = 0$ due to $q(0) \neq 0$). The first condition then finally forces $a = 1$ and hence q is the unique solution of (2.4.28) satisfying the conditions (2.4.33).

In the case $\alpha = 0$ the solution is found to be

$$q(t) = \frac{\sin(\sqrt{2\beta}(\frac{1}{2} - t))}{2\sqrt{2\beta}\cos(\sqrt{\beta/2})} + \frac{1}{2\beta}. \tag{2.4.34}$$

Observe that $q(0)$, the variance of $X(t)$ for any $t \geq 0$, explodes for $\beta \downarrow 0$ and $\beta \uparrow \frac{\pi^2}{2}$. In the first limit the covariance function is nearly constant at a high positive level, in the second limit it resembles a sine function of high amplitude, see Figure 2.4.4 which is however imprecise at the boundaries.

The conditions imposed on the constants automatically have to yield real-valued functions, although the argument relied on complex-valued functions and ω_+, ω_- can be purely imaginary, hence yielding hyperbolic trigonometric functions. Küchler and Mensch (1992) have found similar expressions for the covariance function in the case of a one point delay. Let us summarize.

Proposition 2. *The covariance function q of the stationary solution of the affine SDDE (2.2.10) with exponential weight function (2.4.20) and $v_0 < 0$ is given by*

$$\begin{aligned} q(t) = {} & C_+\xi_+ \sin(\omega_+(|t| - \tfrac{1}{2})) + C_+\eta_+ \cos(\omega_+(|t| - \tfrac{1}{2})) + \\ & C_-\xi_- \sin(\omega_-(|t| - \tfrac{1}{2})) + C_-\eta_- \cos(\omega_-(|t| - \tfrac{1}{2})), \quad t \in [-1,1], \end{aligned}$$

where

- *ω_+ and ω_- are the solutions of (2.4.30) with either nonnegative real or positive imaginary part,*

- *$(\xi_+, \eta_+), (\xi_-, \eta_-)$ span the kernel of the matrix in (2.4.29) for $\omega = \omega_+$ and $\omega = \omega_-$, respectively, and*
- *C_+, C_- are determined from the conditions (2.4.33).*

For $\alpha \geq 0$ the values of ω_+ and ω_- are real. For $\alpha = 0$ we find (2.4.34).

2.5 SDDEs as limits of autoregressive schemes

Suppose the affine SDDE (2.2.10) is discretized according to the Euler scheme with uniform time step $\Delta > 0$. Then the resulting difference equation is an autoregressive scheme:

$$X^\Delta_{n+1} = X^\Delta_n + \Delta \sum_{j=0}^{N} \alpha^\Delta_j X^\Delta_{n-j} + \sqrt{\Delta}\varepsilon_n, \quad \varepsilon_n \sim N(0,1) \ i.i.d. \tag{2.5.35}$$

The coefficients α^Δ_j are chosen such that the point measure $a^\Delta = \sum_{j=0}^N \alpha^\Delta_j \delta_{-j\Delta}$ is – in some later determined sense – close to the original weight measure a. Concerning the notions in connection with autoregressive processes we follow Brockwell and Davis (1996).

In order to justify numerical methods or mathematical modelling with affine SDDEs we are interested in the distribution of the stepwise constant stochastic process

$$Y^\Delta(t) := X^\Delta_{\lfloor \frac{t}{\Delta} \rfloor}, \quad t \geq 0, \tag{2.5.36}$$

in the Skorohod space $\mathcal{D}(\mathbb{R}^+)$. It is shown that for an M^--weight a, small Δ and a reasonable choice of a^Δ stationary solutions X^Δ of the autoregressive scheme exist and that the distributions of the corresponding processes Y^Δ converge to the distribution of the stationary solution X of the affine SDDE in the Skorohod topology for $\Delta \to 0$. With a little bit more effort one can prove the result also for non-Gaussian distributions of (ε_n) with finite second moments. This can be seen as an invariance principle similar to Donsker's theorem on sums of independent random variables.

The restriction to stationary solutions allows us to determine explicitly the (auto)covariance function and spectral density so that a direct comparison of the discrete time and continuous time Gaussian processes is possible. This is the advantage of our method compared to similar results by Scheutzow (1984) who has used Lyapunov functional and Markov chain methods for the investigation of discretized stochastic delay differential equations that might even be nonlinear. Strong (i.e. pathwise) discrete approximations of stochastic delay differential equations for a large class of weight measures have been considered by Hu et al. (2001).

Lemma 4. *If a is an M^--weight and $a^\Delta \xrightarrow{w} a$ holds for $\Delta \to 0$, then the autoregressive scheme (2.5.35) admits a unique stationary solution $(X^\Delta_n)_{n\in\mathbb{N}}$ for small enough Δ, which is causal. In that case the spectral density of X^Δ is given by*

$$f^\Delta(\lambda) = \frac{\Delta}{|1 - e^{-i\lambda} - \Delta \int_{-r}^0 e^{i\lambda s \Delta^{-1}}\, da^\Delta(s)|^2}, \quad \lambda \in [-\pi, \pi].$$

Remark 4. *The assertion of the lemma is to be understood in the sense that for any sequence $\Delta_m \to 0$ with $a^{\Delta_m} \xrightarrow{w} a$ for $m \to \infty$ there is a stationary solution to (2.5.35) for m large enough. This remark also applies to the subsequent statements.*

Proof. According to Brockwell and Davis (1996, p. 83), the characteristic polynomial

$$\varphi^\Delta(z) := 1 - z - \Delta \sum_{j=0}^{N} \alpha_j^\Delta z^j = 1 - z - \Delta \int_{-r}^{0} z^{-s\Delta^{-1}} \, da^\Delta(s), \quad z \in \mathbb{C},$$

has to be studied. If there are no zeros of φ^Δ in the unit circle $\{z \in \mathbb{C} \mid |z| \le 1\}$, then the statement of the lemma follows.

For λ in the strip $\Sigma^\Delta := \{\lambda \in \mathbb{C} \mid \mathrm{Re}(\lambda) \ge 0, |\mathrm{Im}(\lambda)| \le \Delta^{-1}\pi\}$ we consider the function

$$\chi^\Delta(\lambda) := \Delta^{-1}\varphi^\Delta(e^{-\lambda\Delta}) = \frac{1 - e^{-\lambda\Delta}}{\Delta} - \int_{-r}^{0} e^{\lambda s} \, da^\Delta(s). \tag{2.5.37}$$

We set $R := c^{-1} \sup_\Delta \|a^\Delta\|_{TV} < \infty$, where $c > 0$ will be determined later, and split Σ^Δ for small enough Δ into three pieces

$$\begin{aligned} \Sigma_1 &:= \{\lambda \in \mathbb{C} \mid \mathrm{Re}(\lambda) \in [0, R], |\mathrm{Im}(\lambda)| \le R\}, \\ \Sigma_2^\Delta &:= \{\lambda \in \Sigma^\Delta \mid \mathrm{Re}(\lambda) \ge R\}, \\ \Sigma_3^\Delta &:= \{\lambda \in \Sigma^\Delta \mid \mathrm{Re}(\lambda) \le R, |\mathrm{Im}(\lambda)| \ge R\}. \end{aligned}$$

Then $\Delta \to 0$ implies $\chi^\Delta \to \chi_a$ uniformly on the compact rectangle Σ_1, which follows from the argument in the proof of Lemma 2 and the fact that $(1 - e^{-\lambda\Delta})\Delta^{-1} \to \lambda$ holds uniformly in λ on bounded sets.

For $\lambda \in \Sigma_2^\Delta$ we obtain the uniform asymptotic estimate

$$|\chi^\Delta(\lambda)| \ge \frac{|1 - e^{-\lambda\Delta}|}{\Delta} - cR \ge \frac{1 - e^{-R\Delta}}{\Delta} - cR \to (1 - c)R \quad \text{as } \Delta \to 0.$$

For the estimate on Σ_3^Δ we set $f(x) = (1 - \cos(x))x^{-2}$ and $m := \min_{|x| \le \pi} f(x) > 0$. For $\lambda = x + iy \in \Sigma_3^\Delta$ and $\Delta \to 0$ we get

$$\begin{aligned} |\chi^\Delta(\lambda)| &\ge \frac{\sqrt{1 + e^{-2x\Delta} - 2e^{-x\Delta}\cos(y\Delta)}}{\Delta} - cR \\ &\ge \left(\frac{2e^{-x\Delta}(1 - \cos(y\Delta))}{\Delta^2}\right)^{1/2} - cR \\ &\ge \left(\frac{2e^{-R\Delta} m y^2 \Delta^2}{\Delta^2}\right)^{1/2} - cR \\ &\to y\sqrt{2m} - cR. \end{aligned}$$

If we therefore choose $c < \min(1, \sqrt{2m})$, then χ^Δ does not vanish on the strip Σ^Δ for small enough Δ. Consequently, φ^Δ does not vanish on the unit circle due to $\varphi^\Delta(0) = 1$. □

Remark 5. *For later reference we note that the estimate on Σ_3^Δ in the proof yields the first of the inequalities*

$$(\sqrt{2m} - c)y \le |\chi^\Delta(iy)| \le (\sqrt{2M} + c)y, \ \text{ if } y \in \mathbb{R}, |y| \ge R.$$

The second inequality is derived along the same lines, when the triangle inequality instead of the inverse triangle inequality is used and M is set to equal $\max_{|x| \le \pi} f(x)$. Since χ^Δ does not vanish on the imaginary axis, the asymptotic estimate

$$|\chi^\Delta(iy)| \sim (1 + y^2)^{1/2} \tag{2.5.38}$$

for all $y \in \mathbb{R}$ holds uniformly for all small values of Δ.

Proposition 2. *Denote by X^Δ the stationary solution to (2.5.35), which is assumed to exist. Then*

$$\lim_{\Delta\to 0}\sup_{m\in\mathbb{N}_0}|\mathbb{E}[X_0^\Delta X_m^\Delta]-q_a(m\Delta)|=0$$

is satisfied, provided $|\mathcal{F}(a^\Delta-a)(\lambda)|\lesssim(1+\lambda^2)\Delta$ holds for $\lambda\in\mathbb{R}$ and $\Delta\to 0$.

Proof. Using spectral densities, we find

$$\begin{aligned}|\mathbb{E}[X_0^\Delta X_m^\Delta]-q_a(m\Delta)|&=\left|\int_{-\pi}^{\pi}f^\Delta(\lambda)e^{i\lambda m}\,d\lambda-\int_{-\infty}^{\infty}|\chi_a(i\lambda)|^{-2}e^{i\lambda m\Delta}\,d\lambda\right|\\&\le\int_{-\pi/\Delta}^{\pi/\Delta}|\Delta f^\Delta(\Delta\lambda)-|\chi_a(i\lambda)|^{-2}|\,d\lambda \qquad (2.5.39)\\&\quad+\int_{|\lambda|\ge\pi/\Delta}|\chi_a(i\lambda)|^{-2}\,d\lambda.\end{aligned}$$

The last summand tends to zero for $\Delta\to 0$. The integrand in the first summand can be bounded, if the definition of χ^Δ (2.5.37) and the convergence of $\mathcal{F}(a^\Delta-a)$ are used:

$$\begin{aligned}|\Delta f^\Delta(\Delta\lambda)-|\chi_a(i\lambda)|^{-2}|&=||\chi^\Delta(i\lambda)|^{-2}-|\chi_a(i\lambda)|^{-2}|\\&\le\frac{|\chi^\Delta(i\lambda)|+|\chi_a(i\lambda)|}{|\chi^\Delta(i\lambda)\chi_a(i\lambda)|^2}|\chi^\Delta(i\lambda)-\chi_a(i\lambda)|\\&\lesssim\frac{|\chi^\Delta(i\lambda)|+|\chi_a(i\lambda)|}{|\chi^\Delta(i\lambda)\chi_a(i\lambda)|^2}\left(\frac{|1-e^{-i\lambda\Delta}-i\lambda\Delta|}{\Delta}+(1+\lambda^2)\Delta\right).\end{aligned}$$

For real λ with $|\lambda\Delta|\le\pi$ we estimate

$$|e^{-i\lambda\Delta}-1|\sim\lambda\Delta\quad\text{and}\quad e^{-i\lambda\Delta}-1+i\lambda\Delta\sim\lambda^2\Delta^2$$

so that by (2.5.38) a uniform bound of the integrand is obtained:

$$|\Delta f^\Delta(\Delta\lambda)-|\chi_a(i\lambda)|^{-2}|\lesssim(1+\lambda^2)^{-3/2}(\lambda^2\Delta+(1+\lambda^2)\Delta).$$

Thus, the first integral in (2.5.39) also tends to zero as $\Delta\to 0$ because of

$$\int_{-\pi/\Delta}^{\pi/\Delta}(1+\lambda^2)^{-1/2}\Delta\,d\lambda=2\Delta\log(\pi/\Delta+(1+\pi^2/\Delta^2)^{1/2})\to 0.$$

□

Remarks 4.

- *The following choice of the discretisation a^Δ of a satisfies the condition in the preceding proposition. Introduce the distribution function $F_a(x):=a([-r,x])$, $x\in[-r,0]$, of a and put $F^\Delta(x):=F_a(\lfloor\frac{x}{\Delta}\rfloor\Delta)$. Then F^Δ is a right-continuous step-wise constant function, hence a distribution function of a measure a^Δ. Partial integration and the inequality $\|F_a-F^\Delta\|_{L^1}\le\|F_a\|_{BV}\Delta$, which is easily proved for monotonous functions and then extended to bounded variation (BV) functions F_a, yield*

$$\begin{aligned}\left|\int_{-r}^{0}e^{i\lambda s}\,d(a-a^\Delta)(s)\right|&=\left|(F_a-F^\Delta)e^{i\lambda\bullet}|_{-r}^{0}-\int_{-r}^{0}i\lambda e^{i\lambda s}(F_a-F^\Delta)(s)\,ds\right|\\&\le|(F_a-F^\Delta)(-r)e^{-i\lambda r}|+|\lambda|\|F_a-F^\Delta\|_{L^1}\\&\le|(F_a-F^\Delta)(-r)|+\Delta|\lambda|\|F_a\|_{BV([-r,0])}\\&\le\Delta(1+\lambda)\|a\|_{TV}.\end{aligned}$$

- *Concerning the rate of convergence in the proposition it is easily shown that it is at most of order $\Delta\log(\Delta^{-1})$, hence almost linear in Δ. It can hardly be better than linear since $\int_{|\lambda|\geq\pi/\Delta}|\chi_a(i\lambda)|^{-2}\,d\lambda \gtrsim \Delta$ holds. Consider also the next example, where already for $m=0$ only a linear rate of convergence is obtained.*

Example 3. *A discretisation of the Ornstein-Uhlenbeck process with $a=-\delta_0$ leads to the AR(1) process*

$$X_{n+1}^\Delta = (1-\Delta)X_n^\Delta + \sqrt{\Delta}\,\varepsilon_n$$

with covariance function $\mathbb{E}[X_0^\Delta X_m^\Delta] = (2-\Delta)^{-1}(1-\Delta)^m$ of the stationary solution for $\Delta < 1$. The spectral density

$$f_{OU}^\Delta(\lambda) = \frac{\Delta}{|1-e^{-i\lambda}+\Delta|^2}$$

satisfies $\Delta f_{OU}^\Delta(\Delta\lambda) \to \hat{q}_{OU}(\lambda)$ as $\Delta\to 0$.

Theorem 4. *Suppose that a^Δ satisfies the assumptions of Proposition 2 with stationary solution X^Δ of (2.5.35). Then the distributions of Y^Δ from (2.5.36) converge for $\Delta\to 0$ to the distribution of $(X(t),\, t\geq 0)$, the stationary solution to the affine SDDE (2.2.10) with weight a, in the topology of the Skorohod space $\mathcal{D}(\mathbb{R}^+)$. In particular, the distributions converge in $L^\infty([0,T])$ for all $T\geq 0$.*

Proof. Obviously, Y^Δ and X are centered Gaussian processes belonging to $\mathcal{D}(\mathbb{R}^+)$. Therefore the finite dimensional distributions converge if the covariance functions converge pointwise. This follows from regarding

$$c^\Delta(t,s) := \mathbb{E}[Y^\Delta(t)Y^\Delta(s)] = \mathbb{E}[X^\Delta_{\lfloor\frac{t}{\Delta}\rfloor}X^\Delta_{\lfloor\frac{s}{\Delta}\rfloor}],$$

Proposition 2 and the uniform continuity of q_a.

According to Billingsley (1968, Thm. 15.6) the tightness of the distributions of the processes Y^Δ in $\mathcal{D}(\mathbb{R}^+)$ follows from the inequality

$$\mathbb{E}[(Y^\Delta(t_2)-Y^\Delta(t_1))^2(Y^\Delta(t_1)-Y^\Delta(t_0))^2] \lesssim (t_2-t_0)^2 \tag{2.5.40}$$

for all $t_0\leq t_1\leq t_2$. Note that the left hand side vanishes for $t_2-t_0<\Delta$, since then $Y^\Delta(t_1)=Y^\Delta(t_0)$ or $Y^\Delta(t_1)=Y^\Delta(t_2)$ holds. Therefore we can assume $t_2-t_0\geq\Delta$.

Because all random variables in (2.5.40) are Gaussian, the left hand side equals

$$\begin{aligned}(c^\Delta(t_2,t_2)-2c^\Delta(t_2,t_1)+c^\Delta(t_1,t_1))(c^\Delta(t_1,t_1)-2c^\Delta(t_1,t_0)+c^\Delta(t_0,t_0))\\ +2(c^\Delta(t_2,t_1)-c^\Delta(t_2,t_0)-c^\Delta(t_1,t_1)+c^\Delta(t_1,t_0))^2.\end{aligned}$$

We below establish the uniform estimate for all $m,n\in\mathbb{N}_0$

$$|\mathbb{E}[X_0^\Delta X_m^\Delta]-\mathbb{E}[X_0^\Delta X_{m+n}^\Delta]| \lesssim n\Delta, \tag{2.5.41}$$

which shows that the left hand side of (2.5.40) can be bounded up to a constant by

$$4\left(\lfloor\tfrac{t_2}{\Delta}\rfloor-\lfloor\tfrac{t_1}{\Delta}\rfloor\right)\left(\lfloor\tfrac{t_1}{\Delta}\rfloor-\lfloor\tfrac{t_0}{\Delta}\rfloor\right)\Delta^2+2\left(\lfloor\tfrac{t_2}{\Delta}\rfloor-\lfloor\tfrac{t_0}{\Delta}\rfloor\right)^2\Delta^2 \leq 6(t_2-t_0+\Delta)^2.$$

This is the desired bound since $(t_2-t_0+\Delta)\leq 2(t_2-t_0)$ holds for $t_2-t_0\geq\Delta$.

It remains to prove (2.5.41). Put

$$d^\Delta(m) := c^\Delta(0,m\Delta)-c^\Delta_{OU}(0,m\Delta)$$

with c^Δ_{OU} corresponding to the AR(1) process from Example 3. Then by an estimation of the difference of the spectral densities (cf. the properties of χ^Δ in (2.5.37), (2.5.38) and the strategy in the proof of Proposition 1) d^Δ has the property claimed for c^Δ:

$$\begin{aligned}
|d^\Delta(m+n) - d^\Delta(m)| &\leq \int_{-\pi}^{\pi} |f^\Delta(\lambda) - f^\Delta_{OU}(\lambda)||1 - e^{i\lambda n}|\, d\lambda \\
&\leq \int_{-\pi/\Delta}^{\pi/\Delta} ||\chi^\Delta(i\lambda)|^{-2} - |\chi^\Delta_{OU}(i\lambda)|^{-2}||i\lambda\Delta n|\, d\lambda \\
&\leq n\Delta \int_{-\pi/\Delta}^{\pi/\Delta} \frac{|\lambda|(|\chi^\Delta_{OU}(i\lambda)| + |\chi^\Delta(i\lambda)|)}{|\chi^\Delta_{OU}(i\lambda)\chi^\Delta(i\lambda)|^2}(\|a^\Delta\|_{TV} + 1)\, d\lambda \\
&\lesssim n\Delta \int_{-\infty}^{\infty} |\lambda|(1+\lambda^2)^{-3/2}\, d\lambda \\
&\sim n\Delta.
\end{aligned}$$

For $\Delta < 1$ Bernoulli's inequality yields

$$|c^\Delta_{OU}(0,(m+n)\Delta) - c^\Delta_{OU}(0,m\Delta)| = (2-\Delta)^{-1}(1-\Delta)^m(1-(1-\Delta)^n) \leq n\Delta$$

and c^Δ has thus the same asymptotic behaviour as c^Δ_{OU}, which proves the claim (2.5.41). By the definition of the Skorohod topology, the convergence in $L^\infty([0,T])$ is a consequence of the continuity of X. □

Chapter 3

The covariance operator

The study of the covariance operator Q of the stationary solution X on an interval of length r will be essential for all subsequent investigations. Since in the stationary case $(X(t), -r \le t \le 0)$ is a Gaussian process on $C([-r,0])$, the covariance operator maps $M([-r,0])$, the dual of $C([-r,0])$, to $C([-r,0])$. In the first section it is shown that Q is injective. Roughly speaking, Q is an operator that integrates twice. This is given a precise meaning in the second section, where it is shown that Q maps measures with L^2-densities to measures with densities in the Sobolev space $H^2([-r,0])$ for all weight measures a. The Feldman-Hajek theorem is employed to show that the action of the covariance operators Q_a depends in a certain sense continuously on the weight measure a.

These results suffice for the linear estimation theory; for the nonlinear theory however, the action of Q along the scale of Besov spaces has to be studied, which is the content of the last section. Here, the main result is that the operator Q with the Besov space $B^s_{p,\alpha}([-r,0])$ as domain has closed range with codimension 2 in $B^{s+2}_{p,\alpha}([-r,0])$ and that a complementing subspace of the range is given by $\operatorname{span}\{Q\delta_{-r}, Q\delta_0\}$. The essential prerequisite for this property of Q is that the covariance function q is sufficiently regular, which is the case for weight measures in the generalized Besov space $\mathcal{W}^s_{p,\alpha}$.

3.1 Injectivity

The stationary Gaussian process $(X(t), t \ge -r)$ solving the affine SDDE (2.2.10) induces a centred Gaussian distribution on the Banach space $C([-r,0])$ and hence by injection also on the Hilbert space $L^2([-r,0])$. The general theory Vakhaniya et al. (1987, Thm. III.2.2) yields that the covariance operator is a symmetric positive semi-definite operator mapping $M([-r,0])$, the dual of $C([-r,0])$, to $C([-r,0])$ and in the Hilbert space case $L^2([-r,0])$ to $L^2([-r,0])$. The covariance structure can equivalently be described by a symmetric positive semi-definite bilinear form Vakhaniya et al. (1987, Section III.1.1), which leads us to the following definition.

Definition 4. *Let X be the stationary solution of the affine SDDE* (2.2.10) *with weight $a \in M^-$. Then the* covariance operator *$Q_a : M([-r,0]) \to C([-r,0])$ is implicitly defined by the bilinear form*

$$\langle Q_a\mu, \nu\rangle = \mathbb{E}_a\left[\int_{-r}^0 X(t)\, d\mu(t) \int_{-r}^0 X(s)\, d\nu(s)\right], \quad \mu,\nu \in M([-r,0]),$$

where the brackets denote the dual pairing between $C([-r,0])$ and $M([-r,0])$. Equivalently, it can be expressed as integral operator with the covariance function q_a from (2.3.14) as kernel:

$$Q_a\mu(t) = \int_{-r}^{0} q_a(t-s)\,d\mu(s), \quad t \in [-r,0],\ \mu \in M([-r,0]).$$

If there is no danger of ambiguity the subscript a will be dropped. Moreover, if μ has a Lebesgue density f, we shall be inexact and write $Q_a f$ instead of $Q_a\mu$.

A priori, we know that $\langle Q\mu,\mu\rangle \geq 0$ holds for all measures $\mu \in M([-r,0])$. Properties of the covariance operator of the Wiener measure yield even the strict positive definiteness of Q by a change of measure argument.

Proposition 3. *The covariance operator Q is strictly positive definite, i.e. for all $\mu \in M([-r,0])$, $\mu \neq 0$, the inequality $\langle Q\mu,\mu\rangle > 0$ holds.*

Proof. According to Theorem 2 the processes $(X(t),\, 0 \leq t \leq r)$ and $(X(0) + W(t),\, 0 \leq t \leq r)$ induce equivalent measures μ_X and μ_W on $C([0,r])$, for the moment condition use the Fernique theorem Da Prato and Zabczyk (1992, Thm. 2.6). Taking into account that $X(0)$ is a centred Gaussian variable independent of $(W(t),\, t \geq 0)$ and that X is stationary, we obtain for $\mu \in M([-r,0])$

$$\begin{aligned}
\langle Q\mu,\mu\rangle = 0 &\iff \langle X(\bullet + r),\mu\rangle = 0 \quad \mu_X - a.s. \\
&\iff \langle X(0) + W(\bullet + r),\mu\rangle = 0 \quad \mu_W - a.s. \\
&\iff \mathbb{E}[\langle X(0),\mu\rangle^2 + \langle W(\bullet + r),\mu\rangle^2] = 0 \\
&\iff \langle W(\bullet + r),\mu\rangle = 0 \ \mathbb{P}-a.s. \text{ and } \langle \mathbf{1},\mu\rangle = 0.
\end{aligned}$$

Girsanov's theorem implies Revuz and Yor (1999, Cor. VIII.2.3) that the support of the Wiener measure on $C([0,r])$ is the space of all continuous functions starting in zero. The first condition in the last line of equivalences implies that the support of the Wiener measure is contained in the kernel of μ, when μ is regarded as a continuous linear functional. By the support property of the Wiener measure, we thus infer from $\langle Q\mu,\mu\rangle = 0$ that the support of μ is contained in $\{-r\}$. The second condition implies that a one-point measure μ must be identically zero. Thus $\langle Q\mu,\mu\rangle = 0$ implies $\mu = 0$. □

3.2 The domain $L^2([-r,0])$

From now on, we let the integral operator Q act on measures as well as on functions (cf. Definition 4). Then it is shown in this section that Q maps $L^2([-r,0])$ continuously to $H^2([-r,0])$, the Sobolev space of order 2 (cf. Appendix A.1). When writing L^2 or H^s we shall always mean the corresponding function space on $[-r,0]$. The abstract theory Vakhaniya et al. (1987, Thm. IV.2.4) merely yields that Q is a nuclear operator on $L^2([-r,0])$, that is its eigenvalues are summable, whereas our result entails that the eigenvalues are decreasing like n^{-2} (cf. Baumeister (1987) and also the proof of Proposition 9). In the theory of illposed problems such an operator is said to be of smoothing order 2 or to have a degree of ill-posedness 2 Nussbaum and Pereverzev (1999). The fact that Q is not smoothing of higher order is reflected by the closed range of Q in H^2, which means that Q is an isomorphism between L^2 and $Q(L^2)$ in H^2. Similar results for general integral operators have been obtained by Sakhnovich (1996), but by entirely different methods and only for continuous functions.

Example 4. *The covariance operator Q_{OU} of the Ornstein-Uhlenbeck process in Example 2 has the kernel $q_{OU}(t-s) = \frac{1}{2}e^{-|t-s|}$ for $\alpha = -1$. In a distributional sense the identity (with some abuse of notation)*

$$q''_{OU}(t) = \tfrac{1}{2}(e^{-|t|})'' = \tfrac{1}{2}(-\operatorname{sgn}(t)e^{-|t|})' = -\delta_0 + \tfrac{1}{2}e^{-|t|} = -\delta_0 + q_{OU}(t)$$

holds so that for $f \in L^1([-r,0])$ the covariance operator satisfies

$$(Q_{OU}f)''(t) = ((-\delta_0 + q_{OU}) * f)(t) = -f(t) + Q_{OU}f(t) \text{ for a.e. } t \in [-r,0] \;.$$

This shows that the covariance operator Q_{OU} generates the second antiderivative perturbed by an even more regular function.

When function spaces on the whole real line are considered, then the integral operator with kernel q_{OU} is the canonical isomorphism between the Sobolev spaces $H^s(\mathbb{R})$ and $H^{s+2}(\mathbb{R})$, $s \in \mathbb{R}$. The restriction to functions on $[-r,0]$ produces boundary effects: Point measures $\mu = \alpha\delta_{-r} + \beta\delta_0$ with mass at the boundary are mapped by Q_{OU} to the infinitely differentiable function $\alpha q_{OU}(\bullet + r) + \beta q_{OU}$ on $[-r,0]$, since the jumps of the derivative of $Q_{OU}\mu$ at $-r$ ($\alpha \neq 0$) and 0 ($\beta \neq 0$) are "not visible from inside the interval".

This example is a good preparation for the main theorem of this section, but first a lemma is needed.

Lemma 5. *Suppose k is a function in $H^2([0,r])$. Then the integral operator*

$$Kf(t) := \int_{-r}^{0} k(|t-s|)f(s)\,ds, \quad t \in [-r,0],$$

is a continuous linear operator from $L^2([-r,0])$ to $H^2([-r,0])$ with $\|K\|_{L^2\to H^2} \lesssim \|k\|_{H^2}$.

Proof. First consider the following identities in an L^2-sense for $t \in [-r,0]$ and $f \in L^2([-r,0])$

$$\begin{aligned}
(Kf)'(t) &= \int_{-r}^{0} k'(|t-s|)\operatorname{sgn}(t-s)f(s)\,ds,\\
(Kf)''(t) &= \left(\int_{-r}^{\bullet} k'(\bullet - s)f(s)\,ds - \int_{\bullet}^{0} k'(s-\bullet)f(s)\,ds\right)'(t)\\
&= \int_{-r}^{t} k''(t-s)f(s)\,ds + k'(0)f(t) + \int_{t}^{0} k''(s-t)f(s)\,ds + k'(0)f(t)\\
&= \int_{-r}^{0} k''(|t-s|)f(s)\,ds + 2k'(0)f(t).
\end{aligned}$$

By the Cauchy-Schwarz inequality we obtain

$$\begin{aligned}
\|Kf\|_{L^2} &\le 2\|k\|_{L^2}\|f\|_{L^2},\\
\|(Kf)'\|_{L^2} &\le 2\|k'\|_{L^2}\|f\|_{L^2},\\
\|(Kf)''\|_{L^2} &\le 2\|k''\|_{L^2}\|f\|_{L^2} + 2\|k'\|_\infty\|f\|_{L^2}.
\end{aligned}$$

Hence, the Sobolev embedding $H^2 \subset C^1$ proves $\|Kf\|_{H^2} \lesssim \|k\|_{H^2}\|f\|_{L^2}$. □

Theorem 5. *The covariance operator, restricted to $L^2([-r,0])$, is a continuous linear operator $Q_a : L^2([-r,0]) \to H^2([-r,0])$. Its range $Q_a(L^2)$ is closed in $H^2([-r,0])$ with $\operatorname{codim}(Q_a(L^2)) \le 2$.*

The mapping $a \mapsto Q_a$ is weakly continuous in the sense that $a_n \xrightarrow{w} a \in M^-$ implies $\|Q_{a_n} - Q_a\|_{L^2\to H^2} \to 0$.

Proof. Since by Corollary 3 q_a is in $H^2([0,r])$ and symmetric, we can apply the preceding Lemma 5 and deduce the continuity of $Q_a : L^2([-r,0]) \to H^2([-r,0])$. Set $h(t) = q_{OU}(t) - q_a(t)$ with q_{OU} from Example 4. Then from Proposition 1 $h \in H^2([-r,r])$ follows and thus

$$((Q_{OU} - Q_a)f)''(t) = \int_{-r}^{0} h''(t-s)f(s)\,ds, \quad t \in [-r,0],\ f \in L^2([-r,0]),$$

holds. Let us denote the integral operator with kernel h'' by K. Then due to $h'' \in L^2([-r,r])$, the operator K is a Hilbert-Schmidt operator on $L^2([-r,0])$ and therefore compact Dunford and Schwartz (1963, Section XI.6). If D denotes the derivative operator, we obtain the decomposition

$$D^2 Q_a = D^2 Q_{OU} + K = -\operatorname{Id} + Q_{OU} + K.$$

Let $V \subset L^2([-r,0])$ denote the kernel of $D^2 Q_a$ and let W be a complementing subspace of V in $L^2([-r,0])$. By Fredholm theory Zeidler (1995, Section 5.5) the range of $D^2 Q$ is closed and its codimension equals the finite dimension of V. Therefore there exists a complementing subspace U of $D^2 Q(L^2)$ with $\dim U = \dim V$. The situation is illustrated by the following diagram:

$$\begin{array}{ccccc}
L^2 & = & W & \oplus & V \\
 & & & \big\downarrow{\scriptstyle Q} & \\
H^2 & = & Q(L^2) & + & (D^2)^{-1}(U) \\
 & & & \big\downarrow{\scriptstyle D^2} & \\
L^2 & = & D^2 Q(L^2) & \oplus & U
\end{array}$$

While by definition L^2 in the first and in the third line can be written as the direct sum of the respective subspaces, the representation of H^2 in the second line follows from the third line due to $(D^2)^{-1}(D^2 Q(L^2)) = Q(L^2) + \ker(D^2) \subset Q(L^2) + (D^2)^{-1}(U)$. The fact that $Q(V)$ is contained in the kernel of D^2 implies that the operators Q and D^2 each map the vertically corresponding subspaces into each other.

From the injectivity of Q (Proposition 3) $\dim Q(V) = \dim V$ follows. Due to $Q(V) \subset Q(L^2) \cap (D^2)^{-1}(U)$ we obtain

$$\begin{aligned}
\operatorname{codim} Q(L^2) &\le \dim (D^2)^{-1}(U) - \dim Q(V) = (\dim U + \dim \ker(D^2)) - \dim V \\
&= \dim \ker(D^2) = 2.
\end{aligned}$$

Since $Q(L^2)$ has finite codimension in $H^2([-r,0])$, Q has closed range in $H^2([-r,0])$ Zeidler (1995, Ex. 3.12.2).

Suppose that the sequence (a_n) of M^--weights converges weakly to $a \in M^-$. Then the corresponding covariance operators converge due to Lemma 5 and Proposition 1:

$$\|(Q_{a_n} - Q_a)\|_{L^2 \to H^2} \lesssim \|q_{a_n} - q_a\|_{H^2} \to 0.$$

□

As will become apparent in the next section, the codimension of $Q_a(L^2)$ in H^2 is exactly two and a complementing subspace is spanned by $Q_a \delta_{-r}$ and $Q_a \delta_0$.

The fact that the covariance operator Q_a is positive definite allows us to define a positive definite bilinear form, i.e. a scalar product. For the Galerkin method developed in the next chapter it will be essential to establish an ellipticity condition for this scalar product uniformly in the weight a for a varying in a compact set.

Proposition 4. *Let a weight $a_0 \in M^-([-r,0])$ and a weakly compact set $A \subset M^-([-r,0])$ be given. Then the associated covariance operators satisfy*

$$\inf_{a\in A} \inf_{\substack{f\in L^2([-r,0])\\ f\neq 0}} \frac{\langle Q_a f, f\rangle}{\langle Q_{a_0} f, f\rangle} > 0.$$

Proof. Since the covariance operator Q is a positive operator, the bounded operator $Q^{1/2}$ and the unbounded operator $Q^{-1/2}$ are well-defined and positive. In our case the centred Gaussian measures with covariance operator Q_{a_0} and Q_a, $a \in A$, are equivalent on $L^2([-r,0]) \supset C([-r,0])$ by Corollary 2, stationarity and the Fernique theorem. The Feldman-Hajek Theorem Da Prato and Zabczyk (1992, Thm. 2.23) therefore implies that $\operatorname{ran} Q_a^{1/2}$ and $\operatorname{ran} Q_{a_0}^{1/2}$ agree, which – by the closed graph theorem Rudin (1991, Thm. 2.15)– shows that $T_a = Q_{a_0}^{-1/2} Q_a^{1/2}$ is an isomorphism on $L^2([-r,0])$.

According to Theorem 5, Q_a depends in operator norm weakly continuously on a. Therefore, also $Q_a^{1/2}$ Hämarik (1983, p. 75) and thus T_a and its adjoint T_a^* depend in operator norm weakly continuously on a. Finally, the mapping $A \mapsto A^{-1}$ for invertible operators A is norm-continuous Rudin (1991, Thm. 10.11) and thus $a \mapsto \|(T_a^*)^{-1}\|$ is weakly continuous. Hence, observing that $Q_{a_0}^{1/2}$ inherits the injectivity from Q_{a_0}, we obtain by the weak compactness of A

$$\begin{aligned}
\inf_{a\in A} \inf_{\substack{f\in L^2([-r,0])\\ f\neq 0}} \frac{\langle Q_a f, f\rangle}{\langle Q_{a_0} f, f\rangle} &= \inf_{a\in A} \inf_{\substack{h\in Q_{a_0}^{1/2}(L^2)\\ h\neq 0}} \frac{\langle Q_a Q_{a_0}^{-1/2} h, Q_{a_0}^{-1/2} h\rangle}{\langle h, h\rangle} \\
&= \inf_{a\in A} \inf_{h\neq 0} \frac{\langle T_a T_a^* h, h\rangle}{\langle h, h\rangle} \\
&= \inf_{a\in A} \inf_{h\neq 0} \frac{\|T_a^* h\|_{L^2}^2}{\|h\|_{L^2}^2} \\
&= \inf_{a\in A} \|(T_a^*)^{-1}\|^{-2} \\
&> 0.
\end{aligned}$$

In the second line equality holds because the range of the injective and selfadjoint operator $Q_{a_0}^{1/2}$ is dense in $L^2([-r,0])$. □

Remarks 5.

- *It is obvious from the proof that in the preceding proposition the covariance operator of any centred equivalent Gaussian measure can serve as Q_{a_0}. In particular, Q_W defined by*

$$Q_W f(t) := \int_{-r}^{0} (\min(t,s) + r + 1) f(s)\, ds, \quad f \in L^2([-r,0]),\ t \in [-r,0],$$

  which is derived from μ_W in the Girsanov-Theorem 2 with $X(0) \sim N(0,1)$, is eligible.

- *The converse direction of the Feldman-Hajek Theorem gives conditions under which a Gaussian measure $N(\mu, Q)$ on $L^2([-r,0])$ is equivalent to $N(0, Q_W)$. It remains an open and interesting question whether stationary solutions of stochastic differential equations induce these equivalent measures if and only if they solve autonomous affine stochastic differential equations with (possibly unbounded) time delay. Hitsuda (1968) has found a representation of those* *processes X which have an equivalent distribution to ordinary Brownian motion W by means of a stochastic Volterra-type integral:*

$$X(t) = W(t) - \int_0^t \left(\int_0^s k(s,u)\, dW(u) \right) ds, \quad t \in [0,r], \tag{3.2.1}$$

where k is a certain L^2-Volterra kernel. The linear dependence on the past hints towards a positive answer to this question.

3.3 Domains of Besov space type

For the notion of Besov spaces $B^s_{p,\alpha}$ and its properties we refer to Appendix A.2 and the references given there. Just recall the identity $B^s_{2,2} = H^s$ so that the subsequent results are also valid for the scale of L^2-Sobolev spaces. The covariance operator is an integral operator acting on functions defined on the bounded interval $[-r, 0]$. This produces the boundary effect that the point measures δ_{-r} and δ_0 are mapped to functions which are as regular as the covariance function. This will prove to be useful in the section on hypothesis testing as it allows us to distinguish between proper affine SDDEs and equations of Ornstein-Uhlenbeck type. In anticipation of this mapping property we introduce spaces of weights that are linear combinations of weight functions, which have some regularity in the scale of Besov spaces, and of the point measures δ_{-r} and δ_0.

Definition 5. *Let $s \geq 0$ and $p, \alpha \in [1,\infty]$ be given. We introduce the subspace $\mathcal{W}^s_{p,\alpha}$ of weights with λ denoting the Lebesgue measure on $[-r,0]$ by*

$$\mathcal{W}^s_{p,\alpha} := \left\{ \mu_\lambda + m_1\delta_{-r} + m_2\delta_0 \mid \tfrac{d\mu_\lambda}{d\lambda} \in B^s_{p,\alpha}([-r,0]),\, m_1, m_2 \in \mathbb{R} \right\}$$

and equip it with the norm

$$\|\mu\|_{\mathcal{W}^s_{p,\alpha}} := \|\mu\|_{s,p,\alpha} := \|\tfrac{d\mu_\lambda}{d\lambda}\|_{B^s_{p,\alpha}} + |m_1| + |m_2|.$$

We shall abuse notation and write $\mu = f + m_1\delta_{-r} + m_2\delta_0$ with $f = \frac{d\mu_\lambda}{d\lambda}$.

Note that $\mathcal{W}^s_{p,\alpha}$ is canonically isomorphic to the sum of the Banach space $B^s_{p,\alpha}$ and the space $\mathbb{R}^2$. This is the easiest way to see that $\mathcal{W}^s_{p,\alpha}$ is a Banach space. Later on, we shall need the observation that a linear operator T on $\mathcal{W}^s_{p,\alpha}$ satisfies with obvious notation

$$\begin{aligned} \|T\mu\| &= \|T(\mu_\lambda + m_1\delta_{-r} + m_2\delta_0)\| \\ &\leq \|T\mu_\lambda\| + \|T(m_1\delta_{-r} + m_2\delta_0)\| \\ &\leq (\|T|_{B^s_{p,q}}\| + \|T|_{\operatorname{span}(\delta_{-r},\delta_0)}\|)\|\mu\|_{s,p,\alpha}. \end{aligned} \tag{3.3.2}$$

For the investigation of the mapping properties of the covariance operator we first need a result on the regularity of certain convolutions. We have not strived for the outmost generality concerning the range of the indices. Related questions for convolutions between functions from Besov spaces have apparently only been considered on the real axis. If the functions considered on $[-r, 0]$ had compact support in the interior of this interval, we could apply the results from the literature, but here we have to take care of the boundary behaviour. As usual we define $\frac{1}{\infty} := 0$.

Lemma 6. *For functions $g \in B^s_{p,\alpha}([-r,0])$ and $k \in B^{s+1}_{p',\alpha'}([0,r])$, $s \geq 0$ and $1 < p, p' < \infty$, $\alpha \in [p \vee 2, \infty]$, $\alpha' \in [1,2]$ with $\frac{1}{p} + \frac{1}{p'} = \frac{1}{\alpha} + \frac{1}{\alpha'} = 1$, set*

$$L(g,k)(t) := \int_0^t g(u-t)k(u)\, du, \quad t \in [0,r].$$

Then L is a bilinear mapping from $B^s_{p,\alpha}([-r,0]) \times B^{s+1}_{p',\alpha'}([0,r])$ to $B^{s+1}_{p,\alpha}([0,r])$ with

$$\|L(g,k)\|_{s+1,p,\alpha} \lesssim \|g\|_{s,p,\alpha}\|k\|_{s+1,p',\alpha'}.$$

Proof. First, we show for a fixed function g in $B^0_{p,\alpha}([-r,0])$ that $Tk := L(g,k)$ maps $B^s_{p',\alpha'}([0,r])$ to $B^s_{p,\alpha}([0,r])$ for $s \in [0,1]$ and all p and α.

Regarding the case $s = 0$, we obtain by duality and extension to $\mathbb{R}$ Triebel (1983, Thm. 2.11.2, Prop. 3.3.2)

$$\|Tk\|_{L^\infty} \lesssim \sup_{t\in\mathbb{R}} |(g\mathbf{1}_{\mathbb{R}^-}) * (k\mathbf{1}_{\mathbb{R}^+})| \lesssim \|g\|_{0,p,\alpha}\|k\|_{0,p',\alpha'}.$$

Since the L^∞-norm is stronger than the $B^0_{p,\alpha}$-norm for $\alpha \geq p \vee 2$ (Proposition 16 with $m = 0$), we obtain the bound $\|T\|_{B^0_{p',\alpha'} \to B^0_{p,\alpha}} \lesssim \|g\|_{0,p,\alpha}\|k\|_{0,p',\alpha'}$.

For $s = 1$ we consider the derivative of Tk and use the estimate for $s = 0$:

$$\begin{aligned}
\|(Tk)'\|_{0,p,\alpha} &= \left\| \left(\int_{-\bullet}^{0} g(v)k(v+\bullet)\,dv \right)' \right\|_{0,p,\alpha} \\
&= \left\| g(-\bullet)k(0) + \int_0^\bullet g(u-\bullet)k'(u)\,du \right\|_{0,p,\alpha} \\
&\leq \|g\|_{0,p,\alpha}\|k\|_\infty + \|T(k')\|_{0,p,\alpha} \\
&\lesssim \|g\|_{0,p,\alpha}\|k\|_{1,p',\alpha'}.
\end{aligned}$$

We have bounded the $B^1_{p,\alpha}$-seminorm of Tk and hence – using the case $s = 0$ once more – also $\|Tk\|_{1,p,\alpha}$ (cf. Appendix A.2).

Therefore T is a continuous linear operator acting as $T : B^0_{p',\alpha'} \to B^0_{p,\alpha}$ and as $T : B^1_{p',\alpha'} \to B^1_{p,\alpha}$. By the real interpolation theory (Triebel (1983, Thm. 3.3.6) and Bennett and Sharpley (1988, Thm. 5.1.12)) we infer for $s \in [0,1]$ the estimate

$$\|Tk\|_{s,p,\alpha} \lesssim \|g\|_{0,p,\alpha}\|k\|_{s,p',\alpha'}.$$

In the second step, we proceed by an induction argument from s to $s+1$. Suppose $g \in B^s_{p,\alpha}$ and $k \in B^{s+1}_{p',\alpha}$. The weak derivative of $L(g,k)$ is given by (see above)

$$L(g,k)'(t) = g(-t)k(0) + L(g,k')(t), \quad t \in [0,r],$$

which yields for $s \in [0,1]$

$$\|L(g,k)'\|_{s,p,\alpha} \leq \|g\|_{s,p,\alpha}\|k\|_\infty + \|T(k')\|_{s,p,\alpha} \lesssim \|g\|_{s,p,\alpha}\|k\|_{s+1,p',\alpha'}$$

and a fortiori for $s > 1$ by induction

$$\|L(g,k)'\|_{s,p,\alpha} \leq \|g\|_{s,p,\alpha}\|k\|_\infty + \|g\|_{s-1,p,\alpha}\|k'\|_{s,p',\alpha'} \lesssim \|g\|_{s,p,\alpha}\|k\|_{s+1,p',\alpha'}.$$

Since the very first argument provided an estimate for $\|L(g,k)\|_{L^p}$ of the same type, the norm $\|L(g,k)\|_{s+1,p,\alpha}$ is bounded as asserted. □

Proposition 3. *For a weight function $a \in \mathcal{W}^s_{p,\alpha}$ with $v_0(a) < 0$, $s \geq 0$, $1 < p < \infty$ and $\alpha \in [p \vee 2, \infty]$ the covariance function satisfies $q_a \in B^{s+3}_{p,\alpha}([0,r])$.*

Proof. From Corollary 3 we already know that q_a lies in $H^\rho([0,r])$ for all $\rho < \frac{5}{2}$. We now use the fact that q_a satisfies the deterministic delay equation (2.1.1) and is symmetric. For $a = g + \gamma_1\delta_{-r} + \gamma_2\delta_0$ and $t \in [0,r]$ this shows that $q_a''(t)$ equals

$$\begin{aligned}
&\left(\int_{-r}^{0} q_a(\bullet + u)\, da(u)\right)'(t)\\
&= \left(\int_{-r}^{-\bullet} q_a(-\bullet - u)g(u)\, du + \int_{-\bullet}^{0} q_a(\bullet + u)g(u)\, du\right)'(t) + \gamma_1 q_a'(t-r) + \gamma_2 q_a'(t)\\
&= -\int_{-r}^{-t} q_a'(-t-u)g(u)\, du - q_a(0)g(-t) + \int_{-t}^{0} q_a'(t+u)g(u)\, du + q_a(0)g(-t)\\
&\quad + \gamma_1 q_a'(t-r) + \gamma_2 q_a'(t)\\
&= -\int_{0}^{r-t} q_a'(u)g(-u-t)\, du + \int_{0}^{t} q_a'(u)g(u-t)\, du - \gamma_1 q_a'(r-t) + \gamma_2 q_a'(t).
\end{aligned} \tag{3.3.3}$$

We prove the statement $q_a \in B^{s+3}_{p,\alpha}([0,r])$ by setting $\sigma := \sup\{s \geq 0 \,|\, q_a \in B^{s}_{p,\alpha}([0,r])\}$. The bound $\sigma \geq 2$ follows from the embedding $H^\rho \subset B^{\rho-\frac{1}{2}+\frac{1}{p}-\varepsilon}_{p,\alpha}$ (Proposition 16) for all $\varepsilon > 0$ and from $q_a \in H^\rho([0,r])$ for $\rho < \frac{5}{2}$.

By definition of σ the last two summands in (3.3.3) are elements of $B^{\tau_1}_{p,\alpha}$ for all $\tau_1 < \sigma - 1$. By Lemma 6 with $k = q_a'$ the second summand lies in $B^{\tau_2}_{p,\alpha}$ for all τ_2 with $\tau_2 < \sigma - (\frac{2}{p} \vee 1)$ (for $p < 2$ relation (A.2.3) yields $B^{\rho}_{p,\alpha} \subset B^{\rho'}_{p',\alpha'}$ for $\rho' < \rho - \frac{1}{p} + \frac{1}{p'} = \rho + 1 - \frac{2}{p}$) and $\tau_2 \leq s+1$. The substitutions $\tilde{t} := -r - t$ and $\tilde{g}(u) := g(r-u)$ transform the first summand

$$\int_0^{r-t} q_a'(u)g(-u-t)\, du = \int_0^{\tilde{t}} q_a'(u)\tilde{g}(u-\tilde{t})\, du,$$

whence we can conclude by Lemma 6 that this summand also lies in $B^{\tau_2}_{p,\alpha}$ for the same values of τ_2.

This shows that q_a'' is an element of $B^{\tau_1\wedge\tau_2}_{p,\alpha}$, hence for all $\varepsilon \in (0, 2-(\frac{2}{p} \vee 1))$ the inclusion $q_a \in B^{2+\tau_1\wedge\tau_2}_{p,\alpha} \subset B^{2+(\sigma-2+\varepsilon)\wedge(s+1)}_{p,\alpha}$ holds true, from which we conclude $s+1 < \sigma - 2 + \varepsilon$ by the definition of σ and thus $q_a \in B^{s+3}_{p,\alpha}([0,r])$. □

Example 5. *Consider the case $da(s) = -\mathbf{1}_{[-1,0]}(s)ds$ with $r = 2$. Then $v_0(a) < 0$ holds (cf. Lemma 1 or Proposition 1). The choice $r > 1$ is artificial, but serves well as a simple example for effects of a jump. We obtain for $t \in (0,2)$*

$$q_a''(t) = \left(-\int_{-2}^{0} q_a(\bullet + s)\mathbf{1}_{[-1,0]}(s)\, ds\right)'(t) = -q_a(t) + q_a(t-1).$$

By (2.3.18) the derivative of q_a has a jump of size -1 at zero, which yields in our case $q_a'''(1+) - q_a'''(1-) = 1$. On the other hand the third derivative q_a''' exists, is continuous and bounded on the set $[0,2] \setminus \{1\}$.

In terms of function spaces the weight function and the third derivative of the covariance function both lie in $B^s_{p,1}$ for all $s < \frac{1}{p}$ (Appendix A.2), which shows that in this case the covariance function is exactly three times more regular than the weight function.

Theorem 2. *For M^--weights a in $\mathcal{W}^s_{p,\alpha}$, $s \geq 0$, $1 < p < \infty$, $\alpha \in [p \vee 2, \infty]$, the covariance operator is a continuous operator*

$$Q_a : \mathcal{W}^s_{p,\alpha} \to B^{s+2}_{p,\alpha}([-r,0]).$$

Moreover, Q_a is bijective, hence an isomorphism.

Proof. First, let us consider how Q_a acts on functions $f \in B^s_{p,\alpha}$. Since Q_a maps $M([-r,0])$ to $C([-r,0])$ (see the discussion before Definition 4), we only need to estimate $\|(Q_a f)''\|_{s,p,\alpha}$. By symmetry of q_a and by the regularity result $q_a \in B^{s+3}_{p,\alpha}([0,r]) \subset C^2([0,r])$ (Propositions 3 and 16) we obtain for $t \in [-r,0]$ like in the proof of Lemma 5

$$\begin{aligned}(Q_a f)''(t) &= \int_{-r}^{0} q_a''(t-s)f(s)\,ds + 2q_a'(0+)f(t)\\ &= -f(t) + \int_0^{r+t} f(t-u)q_a''(u)\,du + \int_0^{-t} f(t+u)q_a''(u)\,du. \qquad (3.3.4)\end{aligned}$$

Lemma 6 with obvious modifications and the embedding (A.2.3) therefore yield the estimate

$$\|(Q_a f)''\|_{s,p,\alpha} \lesssim \|f\|_{s,p,\alpha} + \|f\|_{s-1,p,\alpha}\|q_a''\|_{s,p',\alpha'} \lesssim (1+\|q_a\|_{s+3,p,\alpha})\|f\|_{s,p,\alpha},$$

which shows that Q_a maps $B^s_{p,\alpha}$ continuously to $B^{s+2}_{p,\alpha}$. Writing the derivative operator as D, we further find for all $\varepsilon \in (0, 2-(\frac{2}{p}\vee 1))$ by (A.2.3)

$$\|(D^2Q_a + \mathrm{Id})f\|_{s+\varepsilon} \lesssim \|f\|_{s+\varepsilon-1,p,\alpha}\|q_a''\|_{s+\varepsilon,p',\alpha'} \lesssim \|q_a\|_{s+3,p,\alpha}\|f\|_{s,p,\alpha}.$$

We infer that $D^2Q_a + \mathrm{Id}$ is a compact operator on $B^s_{p,\alpha}$, because $B^{s+\varepsilon}_{p,\alpha}([-r,0])$ embeds compactly into $B^s_{p,\alpha}([-r,0])$. Exactly the same argument as in the proof of Theorem 5 applies to the Fredholm operator D^2Q_a on $B^s_{p,\alpha}$ and yields that $Q_a(B^s_{p,\alpha})$ is a closed subspace of $B^{s+2}_{p,\alpha}$ of codimension less than or equal to two.

Due to $Q_a\delta_{-r} = q_a(\bullet + r)$ and $Q_a\delta_0 = q_a$ we have $Q_a(\mathrm{span}(\delta_{-r},\delta_0)) \subset B^{s+3}_{p,\alpha} \subset B^{s+2}_{p,\alpha}$ by Proposition 3. It remains to use the injectivity of Q_a on $M([-r,0])$ (Proposition 3) in order to conclude that $Q_a(\mathrm{span}(\delta_{-r},\delta_0))$ is a two-dimensional subspace of $B^{s+2}_{p,\alpha}$ which lies in the complement of $Q_a(B^s_{p,\alpha})$. Since $\operatorname{codim} Q_a(B^s_{p,\alpha}) \le 2$ holds, this codimension must equal two and $Q_a(\mathrm{span}(\delta_{-r},\delta_0))$ is a complementing subspace. This shows that $Q_a : \mathcal{W}^s_{p,\alpha} \to B^{s+2}_{p,\alpha}$ is surjective and hence bijective. Because Q_a is separately continuous on these two subspaces, it is continuous on its span $\mathcal{W}^s_{p,\alpha}$ (cf. (3.3.2)) and by the open mapping theorem it is thus an isomorphism Rudin (1991, Cor. 2.12). □

As in the L^2-case the covariance operator depends continuously on the underlying weight with respect to correctly chosen norms. This will be a corollary to the next proposition, which strengthens the convergence result for the covariance functions from Proposition 1.

Proposition 4. *Suppose $s \ge 0$, $1 < p < \infty$ and $\alpha \in [p \vee 2, \infty]$ are given. If (a_n) is a sequence of $M^- \cap \mathcal{W}^s_{p,\alpha}$-weights that converges in $\mathcal{W}^s_{p,\alpha}$-norm to the $M^- \cap \mathcal{W}^s_{p,\alpha}$-weight a, then $\|q_{a_n} - q_a\|_{B^{s+3}_{p,\alpha}} \to 0$ follows.*

Proof. Put $f_n := q_{a_n} - q_a$ and $a_n = g_n + \gamma_{1,n}\delta_{-r} + \gamma_{2,n}\delta_0$. Similarly as in (3.3.3), the following identities hold for $t \in [0,r]$:

$$\begin{aligned}f_n''(t) &= \left(\int_{-r}^{0} q_{a_n}(\bullet + u)\,da_n(u) - \int_{-r}^{0} q_a(\bullet + u)\,da(u)\right)'(t)\\ &= \left(\int_{-r}^{0} f_n(\bullet + u)\,da_n(u)\right)'(t) + (Q_a(a_n - a)(-\bullet))'(t)\\ &= -\int_{-r}^{-t} f_n'(-t-u)g_n(u)\,du + \int_{-t}^{0} f_n'(t+u)g_n(u)\,du\\ &\quad - \gamma_{1,n}f_n'(r-t) + \gamma_{2,n}f_n'(t) - (Q_a(a_n-a))'(-t)\end{aligned}$$

$$\begin{aligned} &= -\int_0^{r-t} f_n'(u) g_n(-u-t)\,du + \int_{-t}^0 f_n'(u) g_n(u-t)\,du \\ &\quad - \gamma_{1,n} f_n'(r-t) + \gamma_{2,n} f_n'(t) - (Q_a(a_n - a))'(-t). \end{aligned}$$

Therefore we obtain for all $\sigma > 0$ (allowing the value ∞) the estimate

$$\begin{aligned} \|f_n''\|_{\sigma,p,\alpha} &\lesssim \|f_n'\|_{\sigma,p',\alpha'} \|g_n\|_{\sigma-1,p,\alpha} + (|\gamma_{1,n}| + |\gamma_{2,n}|) \|f_n'\|_{\sigma,p,\alpha} \\ &\quad + \|Q_a\|_{B^{(\sigma-1)\vee 0}_{p,\alpha} \to B^{\sigma+1}_{p,\alpha}} \|a_n - a\|_{(\sigma-1)\vee 0,p,\alpha}. \end{aligned} \tag{3.3.5}$$

For $a_n \xrightarrow{w} a$ the covariance functions converge in $H^\rho([0,r])$ for all $\rho < \frac{5}{2}$, whence $\|f_n\|_{\sigma,p,\alpha} \to 0$ holds for all $\sigma < 2 + \frac{1}{p}$. In particular, the convergence $\|f_n\|_{L^p} \to 0$ follows. Hence, the right hand side of estimate (3.3.5) is finite for all $\sigma \in (0, \frac{1}{p})$. Once again using $B^\sigma_{p,\alpha} \subset B^{\sigma-1+\varepsilon}_{p',\alpha'}$ for all $\varepsilon \in (0, 2 - (\frac{2}{p} \vee 1))$, we obtain for all $\sigma \le s+1$

$$\|f_n\|_{\sigma+2,p,\alpha} \lesssim \|f_n\|_{L^p} + \|a_n\|_{s,p,\alpha} \|f_n\|_{\sigma+2-\varepsilon,p,\alpha} + \|Q_a\| \|a_n - a\|_{s,p,\alpha}.$$

Starting with $\sigma_0 = \varepsilon$, we can iterate this estimate $(\sigma_{n+1} := \min(\sigma_n + \varepsilon, s+1))$. Hence $\|f_n\|_{s+3,p,\alpha}$ is bounded by a multiple of $\|f_n\|_{L^p} + \|f_n\|_{2,p,\alpha} + \|a_n - a\|_{s,p,\alpha}$, which tends to zero for $n \to \infty$. This proves $\|f_n\|_{s+3,p,\alpha} \to 0$. □

Remarks 6.

- *In the case of weight functions g_n that converge in H^s-norm, an easier proof can be obtained via the spectral density (cf. Proposition 1), which satisfies*

$$|\hat{q}_{g_n}(\xi) - \hat{q}_g(\xi)| \lesssim (1+\xi^2)^{-3/2} |\hat{g}_n(\xi) - \hat{g}(\xi)|, \quad \xi \in \mathbb{R},$$

  from which it is immediate that $\|g_n - g\|_s \to 0$ implies $\|q_{g_n} - q_g\|_{s+3} \to 0$, even on the whole real line. Unfortunately, there is no simple description of Besov spaces in terms of the Fourier transform. Moreover, the existence of point measures destroys the regularity on the boundary points of the interval $[0,r]$, when the function is regarded on the real line.

- *The case $p \le 1$, which has some importance for results on adaptive approximation Cohen (2000), remains open. For $p \le 1$ we do not have the embedding property $B^s_{p,\alpha} \subset B^{s-1+\varepsilon}_{p',\alpha'}$, which was essential for the induction argument in the proof. The restriction $\alpha \ge p \vee 2$ was only used to guarantee $L^p \subset B^0_{p,\alpha}$. The condition $p < \infty$ was used for the embedding $B^1_{p',\alpha'} \subset C^0$. In any case, we shall need the results only for L^2-Sobolev spaces ($p = 2$, $\alpha = 2$) and Besov spaces with $\alpha = \infty$.*

Corollary 6. *If (a_n) is a sequence of $M^- \cap \mathcal{W}^s_{p,\alpha}$-weights that converges in $\mathcal{W}^\sigma_{p,\alpha}$-norm to the $M^- \cap \mathcal{W}^s_{p,\alpha}$-weight a for some $\sigma > s - 2 + (1 \vee \frac{2}{p})$ and s, p, α as before, then the covariance operators converge in operator norm:*

$$\lim_{n\to\infty} \|Q_{a_n} - Q_a\|_{\mathcal{W}^s_{p,\alpha} \to B^{s+2}_{p,\alpha}} = 0.$$

Proof. From equation (3.3.4) we infer by linearity for $f \in B^s_{p,\alpha}$ and $t \in [-r, 0]$

$$\begin{aligned} &((Q_{a_n} - Q_a)f)''(t) = \\ &\qquad \int_0^{r+t} f(t-u)(q_{a_n} - q_a)''(u)\,du + \int_0^{-t} f(t+u)(q_{a_n} - q_a)''(u)\,du. \end{aligned}$$

By Lemma 6 and by the norm estimates $\|\bullet\|_{s+2,p',\alpha'} \lesssim \|\bullet\|_{\sigma+3,p,\alpha}$ and $\|\bullet\|_{L^{p'}} \lesssim \|\bullet\|_{\sigma+3,p,\alpha}$ we infer the bound

$$\begin{aligned}\frac{\|(Q_{a_n}-Q_a)f\|_{s+2,p,\alpha}}{\|f\|_{s,p,\alpha}} &\lesssim \frac{\|(Q_{a_n}-Q_a)f\|_{L^\infty}+\|f\|_{s-1,p,\alpha}\|(q_{a_n}-q_a)''\|_{s,p',\alpha'}}{\|f\|_{s,p,\alpha}}\\ &\lesssim \|q_{a_n}-q_a\|_{L^{p'}}+\|q_{a_n}-q_a\|_{\sigma+3,p,\alpha}\\ &\lesssim \|q_{a_n}-q_a\|_{\sigma+3,p,\alpha}.\end{aligned}$$

From the preceding Proposition 4 we see that for $\|a_n - a\|_{\sigma,p,\alpha} \to 0$ this bound tends to zero.

Since $(Q_{a_n} - Q_a)\delta_{-r}$ and $(Q_{a_n} - Q_a)\delta_0$ are both expressions of the form $q_{a_n} - q_a$ evaluated on $[0, r]$, Proposition 4 also shows that $Q_{a_n} - Q_a$ tends to zero on $\operatorname{span}(\delta_{-r}, \delta_0)$. By (3.3.2) we therefore conclude $\|Q_{a_n} - Q_a\| \to 0$ on $\mathcal{W}^s_{p,\alpha}$. □

Corollary 7. *For sets $A \subset \mathcal{W}^s_{p,\alpha}$ ($s > 0$ and p, α as before) that are bounded in $\mathcal{W}^s_{p,\alpha}$-norm and satisfy $\sup_{a\in A} v_0(a) < 0$ we obtain*

$$\sup_{a\in A}\|Q_a\|_{\mathcal{W}^s_{p,\alpha}\to B^{s+2}_{p,\alpha}} < \infty, \quad \sup_{a\in A}\|Q_a^{-1}\|_{B^{s+2}_{p,\alpha}\to\mathcal{W}^s_{p,\alpha}} < \infty.$$

Proof. Since A is bounded in $\mathcal{W}^s_{p,\alpha}$, it is by the Besov embeddings relatively compact in any $\mathcal{W}^\sigma_{p,\alpha}$ for $\sigma < s$. Just use the abstract description $\mathcal{W}^\sigma_{p,\alpha} \cong B^\sigma_{p,\alpha} \oplus \delta_{-r} \oplus \delta_0$. Then a compact set K in $B^\sigma_{p,\alpha}$ and a bounded closed (hence compact) set B in $\delta_{-r} \oplus \delta_0$ generate a compact set $K \oplus B$ in $B^\sigma_{p,\alpha} \oplus (\delta_{-r} \oplus \delta_0)$.

In Corollary 6 it was shown that the operator norm of Q_a depends continuously on a in $\mathcal{W}^\sigma_{p,\alpha}$-norm for some $\sigma < s$ due to $p > 1$. Consider the closure $\bar{A}$ of A in $\mathcal{W}^\sigma_{p,\alpha}$, which by Proposition 1 is still contained in M^-. Then $\sup_{a\in\bar{A}}\|Q_a\|$ is finite by the usual compactness argument in $\mathcal{W}^\sigma_{p,\alpha}$.

Finally, the norm continuity of the mapping $Q_a \mapsto Q_a^{-1}$ Rudin (1991, Thm. 10.11) yields the second statement. □

Chapter 4

Convergence results

The law of the solution process of an affine SDDE can be described by an asymptotically sufficient statistic (q_T, b_T). In this chapter the convergence of $\frac{1}{T}q_T$ to q_a for $T \to \infty$ under the stationary law $\mathbb{P}_a$ is established in some strong functional norms. We shall also prove that $\frac{1}{T}b_T$ tends to $Q_a a$, the covariance operator corresponding to the weight a applied to the weight a. These properties are the key for the construction of an estimator of a. The convergence takes place with the usual $\frac{1}{\sqrt{T}}$-rate. Since an argument based on moments, similar to Kolmogorov's continuity theorem, will be applied, the first section deals with bounds on polynomial moments. Another result in this section concerns the convergence of certain exponential moments of wavelet coefficients, which shall be used later for large deviation arguments. The uniform validity of constants over classes of weight measures will become crucial for the statistical results in a minimax setting, but is rather tedious work.

4.1 Moment estimates

In view of the likelihood ratio in Corollary 2 and its simplification in Remark 2 we introduce the random functions q_T and b_T and the random operator Q_T, which are all $\sigma(X(t),\, -r \le t \le T)$-measurable.

Definition 6. *Let $(X(t),\, -r \le t \le T)$ be a stationary solution process of the affine SDDE (2.2.10) with weight $a \in M^-$. We put*

$$\begin{aligned} q_T(u,v) &:= \int_0^T X(t+u)X(t+v)\,dt, \quad u,v \in [-r,0], \\ (Q_T\mu)(s) &:= \int_{-r}^0 q_T(s,v)\,d\mu(v), \quad \mu \in M([-r,0]),\ s \in [-r,0], \\ b_T(s) &:= \int_0^T X(t+s)\,dX(t), \quad s \in [-r,0]. \end{aligned}$$

Remark 6. *Due to the continuity of X it is clear that q_T is a continuous process on $[-r,0]^2$ and Q_T is well defined. With regard to b_T we have the identity*

$$b_T(s) = Q_T a(s) + \int_0^T X(t+s)\,dW(s), \quad s \in [-r,0], \quad \mathbb{P}_a - a.s.$$

We obtain for the stochastic integral by the Burkholder-Davis-Gundy inequality Karatzas and Shreve (1991, Thm. 3.28) and the Lipschitz continuity of the co-

variance function q_a (Corollary 3) uniformly in $s, s' \in [-r, 0]$ for $m \in \mathbb{N}$

$$\mathbb{E}_a \left[\left(\int_0^T X(t+s)\, dW(t) - \int_0^T X(t+s')\, dW(t) \right)^{2m} \right]$$
$$\lesssim \mathbb{E}_a \left[\left(\int_0^T (X(t+s) - X(t+s'))^2\, dt \right)^{m} \right] \lesssim (q_a(0) - q_a(s-s'))^m \lesssim |s-s'|^m.$$

Therefore, we may choose a continuous version of the stochastic integral and hence of b_T by the Kolmogorov continuity theorem Karatzas and Shreve (1991, Thm. 2.8), which we will do from now on.

The ergodicity property of X immediately yields the pointwise results

$$\lim_{T\to\infty} \tfrac{1}{T} q_T(u,v) = q_a(u-v) \qquad \mathbb{P}_a - a.s.,$$
$$\lim_{T\to\infty} \tfrac{1}{T} b_T(s) = \lim_{T\to\infty} \left(\frac{1}{T} Q_T a(s) + \frac{1}{T} \int_0^T X(t+s)\, dW(t) \right) = Q_a a(s) \qquad \mathbb{P}_a - a.s.$$

Together with the injectivity of Q_a this result shows that a is identifiable from an infinitely long continuous observation of a trajectory $(X(t), t \geq -r)$. In order to be able to estimate the risk of an estimator we shall need a more refined result in terms of functional norms. The idea is to apply a version of Kolmogorov's continuity theorem in order to obtain an estimate of $\|\frac{1}{T} q_T(\bullet_1, \bullet_2) - q_a(\bullet_1 - \bullet_2)\|$ in a range of function space norms on $[-r, 0]^2$. For simplicity we shall write $\|\frac{1}{T} q_T - q_a\|$ for this last expression. We now enter into these rather technical estimations. The first one establishes the crucial moment estimate, the second one a simple L^p-estimate and the last two an exponential moment estimate for the wavelet coefficients.

Proposition 5. *With q_T as in Definition 6 and for $u, v, u', v' \in [-r, 0]$, $m \in \mathbb{N}$ and $T > 0$ the estimate*

$$\mathbb{E}_a[(\tfrac{1}{T} q_T(u,v) - q_a(u-v) - \tfrac{1}{T} q_T(u',v') + q_a(u'-v'))^{2m}]$$
$$\lesssim T^{-m}(|u-u'| + |v-v'|)^{2m}$$

holds. For fixed $R > 0$ and $\delta > 0$ the constant may be chosen uniformly for all weights a from $M(R, \delta)$. It then only depends on m.

Proof. The following short hand notations will be used for fixed values u, v, u', v':

$$A(t) := X(t+u), \quad B(t) := X(t+v) - X(t+v'), \quad \alpha := q_a(u-v) - q_a(u-v'),$$
$$C(t) := X(t+v'), \quad D(t) := X(t+u) - X(t+u'), \quad \gamma := q_a(u-v') - q_a(u'-v').$$

We find

$$\tfrac{1}{T} q_T(u,v) - q_a(u-v) - \tfrac{1}{T} q_T(u',v') + q_a(u'-v')$$
$$= \frac{1}{T} \int_0^T (X(t+u)X(t+v) - q_a(u-v) - X(t+u')X(t+v') + q_a(u'-v'))\, dt$$
$$= \frac{1}{T} \int_0^T (A(t)B(t) - \alpha)\, dt + \frac{1}{T} \int_0^T (C(t)D(t) - \gamma)\, dt.$$

Then the Minkowski Inequality in $L^{2m}(\Omega, \mathbb{P})$ yields

$$\mathbb{E}_a \Big[\Big(\tfrac{1}{T} q_T(u,v) - q_a(u-v) - \tfrac{1}{T} q_T(u',v') + q_a(u'-v') \Big)^{2m} \Big] \lesssim$$
$$\mathbb{E}_a \left[\left(\frac{1}{T} \int_0^T (A(t)B(t) - \alpha)\, dt \right)^{2m} \right] + \mathbb{E}_a \left[\left(\frac{1}{T} \int_0^T (C(t)D(t) - \alpha)\, dt \right)^{2m} \right].$$

Both summands can be estimated in exactly the same way, so we only present the estimation of the first summand.

With the abbreviations $A_i := A(t_i)$, $B_i := B(t_i)$ we are lead to

$$\begin{aligned}
&\mathbb{E}_a\left[\left(\frac{1}{T}\int_0^T (A(t)B(t)-\alpha)\,dt\right)^{2m}\right] \\
&= \frac{1}{T^{2m}}\int_0^T\cdots\int_0^T \mathbb{E}_a\left[\prod_{i=1}^{2m}(A_iB_i-\alpha)\right] dt_{2m}\dots dt_1 \\
&= \frac{(2m)!}{T^{2m}}\int_0^T\int_0^{t_1}\cdots\int_0^{t_{2m-1}} \mathbb{E}_a\left[\prod_{i=1}^{2m}(A_iB_i-\alpha)\right] dt_{2m}\dots dt_1.
\end{aligned}$$

The application of the Fubini Theorem is justified by the Fernique theorem Da Prato and Zabczyk (1992, Thm. 2.6) for the Gaussian measure induced by X on $C([-r,T])$.

In order to evaluate the expected value of the product, let us introduce the family $P_2(2n)$ of all partitions of the set $\{1,\dots,2n\}$ into subsets with two elements. An easy argument based on the characteristic function shows that for a centered Gaussian random vector $(N_1,\dots,N_{2n})$ the formula

$$\mathbb{E}\left[\prod_{i=1}^{2n} N_i\right] = \sum_{\Gamma\in P_2(2n)}\prod_{(k,l)\in\Gamma}\mathbb{E}[N_kN_l]$$

is valid. In our case we choose $n = 2m$, $N_{2i-1} = A_i$, $N_{2i} = B_i$. Since $\alpha = \mathbb{E}[N_{2i-1}N_{2i}]$ holds, terms involving neighbouring random variables N_{2i-1}, N_{2i} cancel (proof by induction over n) so that

$$\mathbb{E}_a\left[\prod_{i=1}^{2m}(A_iB_i-\alpha)\right] = \sum_{\substack{\Gamma\in P_2(4m)\\ \forall i:\,(2i-1,2i)\notin\Gamma}}\prod_{(k,l)\in\Gamma}\mathbb{E}_a[N_kN_l] \tag{4.1.1}$$

holds true. When the definition of A_i and B_i is resubstituted, then the right hand side is an expression in terms of the covariance function q_a. From Corollary 3 we use the Lipschitz continuity of q_a on $\mathbb{R}$ and of q_a' on the real line away from zero to obtain for $t_k \geq t_l$ and $\delta < -v_0(a)$ with uniform constants

$$\begin{aligned}
|\mathbb{E}_a[A_kA_l]| &= |q_a(t_k-t_l)| \lesssim e^{-\delta(t_k-t_l)},\\
|\mathbb{E}_a[A_kB_l]| &= |q_a(t_k-t_l+u-v) - q_a(t_k-t_l+u-v')|\\
&\lesssim e^{-\delta(t_k-t_l-r)}|v-v'|,\\
|\mathbb{E}_a[B_kB_l]| &= |2q_a(t_k-t_l) - q_a(t_k-t_l+v-v') - q_a(t_k-t_l+v'-v)|\\
&\lesssim \min\Big(4e^{-\delta(t_k-t_l-r)},\, 2|v-v'|\mathbf{1}_{[0,|v-v'|]}(t_k-t_l)+\\
&\quad + \int_0^{|v-v'|}|q_a'(t_k-t_l+h) - q_a'(t_k-t_l-h)|\,dh\,\mathbf{1}_{(|v-v'|,\infty)}(t_k-t_l)\Big)\\
&\lesssim \min(e^{-\delta(t_k-t_l-r)},\, |v-v'|(\mathbf{1}_{[0,|v-v'|]}(t_k-t_l)+|v-v'|).
\end{aligned}$$

Every product on the right of (4.1.1) consists of $2m$ factors, at least m of which involve a covariance formed with some B_k. The bounds on the covariances, which have just been derived, decrease with increasing distance $t_k - t_l$ so that we can estimate for all partitions Γ

$$\left|\prod_{(k,l)\in\Gamma}\mathbb{E}[N_kN_l]\right| \lesssim \prod_{i=1}^{m} e^{-\delta(t_{2i-1}-t_{2i}-r)}|v-v'|(\mathbf{1}_{[0,|v-v'|]}(t_{2i-1}-t_{2i})+|v-v'|).$$

The gain of this estimation argument lies in the fact that we can perform the multiple integration by a pairwise evaluation:

$$\begin{aligned}
&\mathbb{E}_a\left[\left(\frac{1}{T}\int_0^T A(t)B(t)-\alpha\,dt\right)^{2m}\right]\\
&\quad\lesssim \frac{(2m)!|P_2(4m)|}{T^{2m}}\int_0^T dt_1\int_0^{t_1} dt_2\cdots\int_0^{t_{2m-1}} dt_{2m}\\
&\quad\quad |v-v'|^m\prod_{i=1}^m e^{-\delta(t_{2i-1}-t_{2i})}(\mathbf{1}_{t_{2i-1}-t_{2i}\le|v-v'|}+|v-v'|)\\
&\quad\lesssim \frac{|v-v'|^m}{T^{2m}}\left(\int_0^T\int_0^t e^{-\delta s}(\mathbf{1}_{s\le|v-v'|}+|v-v'|)\,ds\,dt\right)^m\\
&\quad\lesssim T^{-m}|v-v'|^{2m}.
\end{aligned}$$

Note that the constants in this calculation depend on m, δ and a uniform constant for $a\in M(R,\delta+\varepsilon)$ with $\varepsilon>0$ any constant. The assertion then follows from using $\delta-\varepsilon$ instead of δ. ☐

Lemma 7. *The random function q_T satisfies*

$$\mathbb{E}_a[\|\tfrac{1}{T}q_T-q_a\|_{L^{2m}}^{2m}]\lesssim T^{-m},$$

involving a constant which may be chosen uniformly for all a from $M(R,\delta)$ for fixed $R>0$ and $\delta>0$.

Proof. Exactly as in the proof of Proposition 5 (use $|\mathbb{E}_a[X(t_i+u)X(t_j+v)]|\lesssim e^{-\delta(|t_i-t_j|-|u-v|)}$) we obtain

$$\begin{aligned}
\mathbb{E}_a[\|\tfrac{1}{T}q_T-q_a\|_{L^{2m}}^{2m}] &= \int_{-r}^0 du\int_{-r}^0 dv\,T^{-2m}\int_0^T dt_1\cdots\int_0^T dt_{2m}\\
&\qquad \mathbb{E}_a\left[\prod_{i=1}^{2m}(X(t_i+u)X(t_i+v)-q_a(u-v))\right]\\
&\lesssim \sup_{u,v}T^{-2m}\int_0^T\cdots\int_0^{t_{2m-1}}\prod_{i=1}^m e^{-\delta(t_{2i-1}-t_{2i}-|u-v|)}\,dt_{2m}\dots dt_1\\
&\lesssim T^{-m}.
\end{aligned}$$

The uniformity of the constant follows again from Corollary 3. ☐

The next proposition and the subsequent lemma will prove to be useful for the wavelet thresholding algorithm in the chapter about adaptive estimation and for the testing problem. It is of no importance for the linear estimation theory. For the notions regarding wavelets the appendix A.3 should be consulted, in particular the notion of regularity of a wavelet basis here has a slightly different meaning than usually.

Proposition 6. *Let (ψ_λ) be a compactly supported 2-regular wavelet basis of $L^2([-r,0])$. Then there exists a constant $K>0$ such that for $\mu\in M([-r,0])$*

$$\sup_\lambda \mathbb{E}_a[\cosh(\alpha T^{1/2}2^{3|\lambda|/2}\langle(\tfrac{1}{T}Q_T-Q_a)\mu,\psi_\lambda\rangle)]\le\frac{1}{1-K^2\alpha^2\|\mu\|_{TV}^2}$$

holds for all $\alpha\in[0,(K\|\mu\|_{TV})^{-1})$. For fixed $R>0$, $\delta>0$ the constant K may be chosen uniformly for all weights a from $M(R,\delta)$.

In addition, we obtain for $m \in \mathbb{N}$

$$\sup_\lambda \mathbb{E}_a[\langle(\tfrac{1}{T}Q_T - Q_a)\psi_\lambda, \psi_\lambda\rangle^{2m}] \lesssim T^{-m}2^{-4|\lambda|m}$$

with a constant that may equally be chosen uniformly over $M(R,\delta)$.

Proof. Due to $\cosh(x) = \sum_m \frac{x^{2m}}{(2m)!}$ we shall again estimate polynomial moments, but this time the dependence on m becomes crucial. The statement concerning polynomial moments will be a simple by-product during the course of the proof.

Recall that $\langle \bullet, \bullet\rangle$ denotes the L^2-scalar product as well as the dual pairing between continuous functions and measures such that $\langle K\mu, f\rangle = \langle \mu, Kf\rangle$ holds for integral operators K with a symmetric continuous kernel, continuous functions f and finite measures μ. Using the finiteness of $\mathbb{E}[\|X\|_{C([-r,T])}]$ by the Fernique theorem on $C([-r,T])$ as requirement for the Fubini Theorem we obtain:

$$\begin{aligned}
&\mathbb{E}_a[\langle(\tfrac{1}{T}Q_T - Q_a)\mu, \psi_\lambda\rangle^{2m}] \\
&= \mathbb{E}_a[\langle \mu, (\tfrac{1}{T}Q_T - Q_a)\psi_\lambda\rangle^{2m}] \\
&= \int_{[-r,0]^{2m}} \mathbb{E}_a\left[\prod_{i=1}^{2m}(\tfrac{1}{T}Q_T - Q_a)\psi(u_i)\right] d\mu(u_{2m})\dots d\mu(u_1) \\
&\le \|\mu\|_{TV}^{2m} \sup_{u_j} \int_{[-r,0]^{2m}} \mathbb{E}_a\left[\prod_{i=1}^{2m}(\tfrac{1}{T}q_T(u_i,v_i) - q_a(u_i,v_i))\right]\prod_{i=1}^{2m}\psi_\lambda(v_i)\,dv_{2m}\dots dv_1 \\
&= \|\mu\|_{TV}^{2m} T^{-2m} \sup_{u_j}\int_{[-r,0]^{2m}} dv_{2m}\dots dv_1 \int_{[0,T]^{2m}} dt_{2m}\dots dt_1 \\
&\quad \mathbb{E}_a\left[\prod_{i=1}^{2m}(X(t_i+u_i)X(t_i+v_i) - q_a(u_i - v_i))\right]\prod_{i=1}^{2m}\psi_\lambda(v_i).
\end{aligned}$$

As in (4.1.1) the expected value of the product equals

$$\sum_{\substack{\Gamma\in P_2(4m)\\ \forall i:(2i-1,2i)\notin\Gamma}} \prod_{(k,l)\in\Gamma} q_a(z_k - z_l)$$

with $z_{2i-1} = t_i + u_i$ and $z_{2i} = t_i + v_i$. Changing the order of integration (q_a is continuous), we start with the integration over v_i, $i = 1,\dots,2m$. Since any v_i appears only once in the product, we have to deal with products over terms which have one of the following three forms:

$$\begin{aligned}
&q_a(t_i + u_i - t_j - u_j), && \text{(I)},\\
&\int_{-r}^0 q_a(t_i+u_i-t_j-v_j)\psi_\lambda(v_j)\,dv_j && \text{(II)},\\
&\int_{-r}^0\int_{-r}^0 q_a(t_i+v_i-t_j-v_j)\psi_\lambda(v_i)\psi_\lambda(v_j)\,dv_i\,dv_j && \text{(III)}.
\end{aligned}$$

For the factor (I) we shall use $|q_a(t_i+u_i-t_j-u_j)| \le C_1 e^{-\delta|t_i-t_j|}$ derived from (2.1.7) or Corollary 3 for $\delta < -v_0(a)$.

The Lipschitz constant of $q_a(t_i+u_i-t_j-\bullet)$ on $[-r,0]$ is of order $e^{-\delta(|t_i-t_j|-r)}$ by Corollary 3, which implies the existence of a constant C_2 such that the modulus of the integral (II) is smaller than $C_2 2^{-3|\lambda|/2}e^{-\delta|t_i-t_j|}$ (Corollary 13).

For the estimation of the integral (III) we let S denote the length of an interval supporting ψ and distinguish the cases (1) $|t_i-t_j| > 2^{-|\lambda|}S$ and (2) $|t_i-t_j| \le 2^{-|\lambda|}S$.

A substitution gives

$$\int_{-r}^{0}\int_{-r}^{0} q_a(t_i+v_i-t_j-v_j)\psi_\lambda(v_i)\psi_\lambda(v_j)\,dv_i\,dv_j$$
$$= \iint_{|v_i-v_j|\le S} q_a(t_i-t_j+2^{-|\lambda|}(v_i-v_j))2^{-|\lambda|}\psi(v_i)\psi(v_j)\,dv_i\,dv_j,$$

which shows that in case (1) q_a needs only to be evaluated at either positive arguments or at negative ones. Due to the Lipschitz continuity of q_a' with exponentially decaying norm (Corollary 3) the estimate in Corollary 13 shows that the modulus of (III) is in case (1) smaller than $C_3 2^{-3|\lambda|}e^{-\delta|t_i-t_j|}$, $C_3>0$ a constant. In case (2) q_a is at least Lipschitz continuous and the modulus of (III) is by the same arguments smaller than $C_4 2^{-2|\lambda|}e^{-\delta|t_i-t_j|}$, $C_4>0$ a constant.

Finally note that each u_i and v_i appears exactly once in the product and that each t_i appears twice so that with $C:=\max_j C_j$

$$\int_{[-r,0]^{2m}} dv_{2m}\dots dv_1\,\mathbb{E}_a\left[\prod_{i=1}^{2m}(X(t_i+u_i)X(t_i+v_i)-q_a(u_i-v_i))\right]\prod_{i=1}^{2m}\psi_\lambda(v_i)$$
$$\le \sum_{\Gamma} 2^{-3|\lambda|m}C^{2m}\prod_{(k,l)\in\Gamma}\big(1+2^{|\lambda|}\mathbf{1}_{\{k,l\text{ even},|t_{k/2}-t_{l/2}|\le S2^{-|\lambda|}\}}\big)e^{-\delta|t_{\lceil k/2\rceil}-t_{\lceil l/2\rceil}|}.$$

If we now integrate over t_i, $i=1,\dots,2m$, and use the symmetry in $t_1,\dots,t_{2m}$, then we arrive at

$$\begin{aligned}
&\mathbb{E}_a[\langle(\tfrac{1}{T}Q_T-Q_a)\mu,\psi_\lambda\rangle^{2m}]\\
&\quad\le \|\mu\|_{TV}^{2m}T^{-2m}2^{-3|\lambda|m}C^{2m}|P_2(4m)|\\
&\qquad\int_{[0,T]^{2m}}\prod_{i=1}^{m}\big(1+2^{|\lambda|}\mathbf{1}_{|t_{2i-1}-t_{2i}|\le S2^{-|\lambda|}}\big)e^{-\delta|t_{2i-1}-t_{2i}|}\,dt_{2m}\dots dt_1\\
&\quad= \|\mu\|_{TV}^{2m}T^{-2m}2^{-3|\lambda|m}C^{2m}|P_2(4m)|\\
&\qquad\left(\int_0^T\int_{-s}^{T-s}\big(1+2^{|\lambda|}\mathbf{1}_{|t|\le S2^{-|\lambda|}}\big)e^{-\delta|t|}\,dt\,ds\right)^m\\
&\quad\le \|\mu\|_{TV}^{2m}T^{-m}2^{-3|\lambda|m}C^{2m}|P_2(4m)|(\tfrac{2}{\delta}+2S)^m.
\end{aligned}$$

In particular, putting $\mu=\psi_\lambda$ such that $\|\mu\|_{TV}=\|\psi_\lambda\|_{L^1}\sim 2^{-|\lambda|/2}$ holds, we have proved the additional statement on polynomial moments.

Since $|P_2(4m)|\le 2^{2m}(2m)!$ holds (it equals the standard Gaussian moments), we finally obtain for sufficiently small $\alpha>0$ the uniform estimate in λ

$$\mathbb{E}_a[\cosh(\alpha T^{1/2}2^{3|\lambda|/2}\langle(\tfrac{1}{T}Q_T-Q_a)\mu,\psi_\lambda\rangle)]\le\sum_{m=0}^{\infty}2^{2m}\alpha^{2m}\|\mu\|_{TV}^{2m}C^{2m}(2+2\delta S)^m$$
$$=\frac{1}{1-4\alpha^2\|\mu\|_{TV}^2C^2(2+2\delta S)}.$$

Furthermore, the deduction relied only on Corollary 3 concerning the dependence on a whence the uniformity of the constant follows. Finally, setting $K:=2C\sqrt{2+2\delta S}$ finishes the proof. □

Lemma 8. *Let (ψ_λ) be a compactly supported and 1-regular wavelet basis of $L^2([-r,0])$. Then there exists a constant $K>0$ such that the estimate*

$$\sup_\lambda \mathbb{E}_a[\exp(\beta T^{-1/2}2^{|\lambda|}|\langle b_T-Q_Ta,\psi_\lambda\rangle|)]\le\frac{1}{1-\beta^2K^2}<\infty$$

is satisfied for $\beta < K^{-1}$. For fixed $R > 0$, $\delta > 0$ the constant K may be chosen uniformly for all weights a from $M(R, \delta)$.

Proof. By the trivial inequality $e^{|x|} \le e^x + e^{-x}$, $x \in \mathbb{R}$, it suffices to consider

$$\mathbb{E}_a[\exp(\pm C\langle b_T - Q_T a, \psi_\lambda\rangle)]$$

with $C = \beta T^{-1/2} 2^{|\lambda|}$. The Fubini theorem for stochastic integrals Protter (1992, Thm. 46) thus yields equality with

$$\mathbb{E}_a\left[\exp\left(\pm C \int_0^T \int_{-r}^0 X(t+s)\psi_\lambda(s)\,ds\,dW(t)\right)\right].$$

Regarding C as fixed for the moment, we know by the Itô formula that

$$\exp\left(\pm C \int_0^T \int_{-r}^0 X(t+s)\psi_\lambda(s)\,ds\,dW(t) - \frac{C^2}{2}\int_0^T \left(\int_{-r}^0 X(t+s)\psi_\lambda(s)\,ds\right)^2 dt\right)$$

is a nonnegative local martingale in T with respect to the filtration $(\mathcal{F}_T)$, hence by Fatou's lemma it is a supermartingale and its expected value is bounded by one. For any C we therefore obtain by the Cauchy-Schwarz inequality

$$\mathbb{E}_a[\exp(\pm C\langle b_T - Q_T a, \psi_\lambda\rangle)] \le \mathbb{E}_a\left[\exp\left(2\frac{C^2}{2}\int_0^T \left(\int_{-r}^0 X(t+s)\psi_\lambda(s)\,ds\right)^2 dt\right)\right].$$

The right-hand side is easy to evaluate since X is a stationary Gaussian process. Jensen's inequality and an evaluation of Gaussian moments give

$$\begin{aligned}
\mathbb{E}_a\left[\left(C^2\int_0^T \langle X(t+\bullet), \psi_\lambda\rangle^2\,dt\right)^m\right] &\le C^{2m}T^m \frac{1}{T}\int_0^T \mathbb{E}_a[\langle X(t+\bullet), \psi_\lambda\rangle^{2m}] \\
&= C^{2m}T^m \langle Q_a\psi_\lambda, \psi_\lambda\rangle^m \frac{(2m)!}{2^m m!} \\
&\le \left(2C^2 T\langle Q_a\psi_\lambda, \psi_\lambda\rangle\right)^m m!
\end{aligned}$$

and Corollary 13 shows with a constant $K > 0$ for $C > 0$ sufficiently small

$$\begin{aligned}
\mathbb{E}_a\left[\exp\left(C^2\int_0^T \left(\int_{-r}^0 X(t+s)\psi_\lambda(s)\,ds\right)^2 dt\right)\right] &\le \sum_{m=0}^\infty (2C^2T\langle Q_a\psi_\lambda, \psi_\lambda\rangle)^m \\
&= \frac{1}{1 - 2C^2T\langle Q_a\psi_\lambda, \psi_\lambda\rangle} \\
&\le \frac{1}{1-\beta^2K^2}.
\end{aligned}$$

By the Corollaries 3 and 13 we conclude that the estimate holds for all $\beta < K^{-1}$ and that we can choose K uniformly for $a \in M(R, \delta)$. □

4.2 Convergence in function spaces

By Kolmogorov's continuity theorem or just by partial integration in (2.3.13) we see that the trajectories $(X(t), -r \le t \le T)$ of a stationary solution lie in the Hölder space $C^\alpha([-r, T])$ for $\alpha < \frac{1}{2}$, i.e. they are as regular as Brownian motion. On the other hand, the function q_T lies in $C^\alpha([-r, 0]^2)$ for $\alpha < 1$. We shall prove in this section that $\frac{1}{T}q_T$ converges to q_a in C^α-norm for all $\alpha < 1$. Note also that by Lemma 6 with $\mu = \delta_u$, $u \in [-r, 0]$, the wavelet coefficients of $\frac{1}{T}q_T(u, \bullet) - q_a(u - \bullet)$ are in the mean of order $2^{-3|\lambda|/2}$. If we could exchange the order of taking supremum and expected value, this would imply that q_T would even be Lipschitz continuous. This stronger continuity property can in general not be expected due to $X \notin C^{1/2}([0, T])$ and the convolution-type integral in the definition of q_T.

Proposition 7. *For any $\alpha < 1$ and $1 \le p < \infty$ the estimate*

$$\mathbb{E}_a[\|\tfrac{1}{T}q_T(\bullet_1, \bullet_2) - q_a(\bullet_1 - \bullet_2)\|^p_{C^\alpha([-r,0]^2)}]^{1/p} \lesssim T^{-1/2}$$

holds with a constant that for weights $a \in M(R, \delta)$ with $R > 0$, $\delta > 0$ can be chosen uniformly and is independent of T.

Proof. Using the inner description of Sobolev spaces $W^{\sigma,p}([-r, 0]^2)$ (Appendix A.1), we obtain from Lemma 7 and Proposition 5 with uniform constants for $\sigma < 1$ and $m \in \mathbb{N}$ such that $m\sigma \le m - 1$

$$\begin{aligned}
\mathbb{E}_a[\|\tfrac{1}{T}q_T - q_a\|^{2m}_{W^{\sigma,2m}}] &\sim \mathbb{E}_a[\|\tfrac{1}{T}q_T - q_a\|^{2m}_{L^{2m}}] + \\
&\quad \int_{[-r,0]^2}\int_{[-r,0]^2} \frac{\mathbb{E}_a[((\frac{1}{T}q_T - q_a)(x) - (\frac{1}{T}q_T - q_a)(y))^{2m}]}{|x-y|^{2m\sigma+2}}\,dx\,dy \\
&\lesssim T^{-m} + T^{-m}\int_{[-r,0]^2}\int_{[-r,0]^2} |x-y|^{2m-(2m\sigma+2)}\,dx\,dy \\
&\sim T^{-m}.
\end{aligned}$$

An application of the Sobolev embedding theorem (Appendix A.1) then proves for $\alpha < \sigma - \frac{2}{2m}$

$$\mathbb{E}_a[\|\tfrac{1}{T}q_T - q_a\|^{2m}_{C^\alpha}]^{1/2m} \lesssim T^{-1/2}.$$

By choosing $\sigma > \alpha$ and m sufficiently large the assertion follows. □

By Remark 6 b_T has a version with α-Hölder continuous trajectories, $\alpha < \frac{1}{2}$, and from Lemma 8 and Proposition 7 one can even derive convergence in H^α-norm using wavelets. We shall however be content with a convergence result merely in L^2.

Corollary 8. *The estimate*

$$\mathbb{E}_a[\|\tfrac{1}{T}b_T - Q_a a\|^2_{L^2([-r,0])}]^{1/2} \lesssim T^{-1/2}$$

holds with a constant that for weights $a \in M(R, \delta)$ with $R > 0$, $\delta > 0$ can be chosen uniformly and is independent of T.

Proof. Applying the Fubini-Tonelli Theorem for positive integrands, the SDDE (2.2.10), an easy norm estimate and Proposition 7, we obtain

$$\begin{aligned}
\mathbb{E}_a[\|\tfrac{1}{T}b_T - Q_a a\|^2_{L^2}] &\le 2\int_{-r}^0 \mathbb{E}_a\Bigg[\left(\frac{1}{T}\int_0^T X(t+s)\,dW(t)\right)^2 \\
&\quad + \left(\frac{1}{T}\int_0^T X(t+s)\int_{-r}^0 X(t+u)\,da(u)\,dt - Q_a a(s)\right)^2\Bigg]\,ds \\
&\le 2\int_{-r}^0 T^{-1}q_a(0)\,ds + 2r\|a\|_{TV}\,\mathbb{E}_a[\|\tfrac{1}{T}q_T - q_a\|^2_{C([-r,0]^2)}] \\
&\lesssim T^{-1}.
\end{aligned}$$

The uniformity of the constants follows from Propositions 1 and 7. □

Chapter 5

The Galerkin estimator

In the preceding chapter the convergence of $\frac{1}{T}q_T$ to q_a and of $\frac{1}{T}b_T$ to $Q_a a$ was established. Abstractly speaking, we can infer from the observations an operator A_η close to another operator A and data y_δ close to y and we want to determine from this a good approximation for $A^{-1}y$. This leads to an ill-posed problem because the operator A is injective, but its inverse is not continuous (Chapter 3) so that small errors in the data can lead to arbitrarily large estimation errors. An additional difficulty lies in the fact that we only know an approximation A_η of the operator A.

In the first section, an introduction to the deterministic Galerkin method for such ill-posed problems is given. This method is used in the second section for the construction of an estimator of the weight function. For classes of regular weight functions a uniform asymptotic upper bound for this estimator is proved in the third section. This is followed by an investigation of the behaviour of the estimator for discrete time observations of maximal distance Δ. Asymptotic risk bounds are obtained for $T \to \infty$ and $\Delta \to 0$. A discussion of the case where the estimator is applied to a process with a general – not necessarily absolutely continuous – weight measure concludes the chapter.

5.1 Inverse problems

There are several methods to treat ill-posed inverse problems. The general paradigm is that the problem P is approximated by a scale of well-posed inverse problems P_h which, as h becomes smaller, approximate the original problem P better, but become more ill-conditioned. Under a compactness assumption on the unknown true solution it is possible to calibrate the approximation error ("the bias term") and the bad condition ("the variance term") by a right choice of h depending on the error level, such that the solution of P_h converges to the true solution of P as the error level tends to zero. A good introduction to these concepts is given by Baumeister (1987).

Among the classical methods for ill-posed problems the so-called Galerkin, Rayleigh-Ritz or variational method is particularly well suited for our problem because it is designed for positive definite operators, its numerical implementation is straightforward and it can easily be extended to approximately known operators. We shall be working with an L^2-type risk function so that Hilbert space techniques can be employed.

Given the identity $Ax = y$ in a Hilbert space H, where A is a selfadjoint and positive definite linear operator on H, x and y are elements of H, the abstract ill-posed inverse problem consists of finding a good approximation x_n of x from the knowledge of the approximations y_δ of y and A_η of A. The approach of the Galerkin method is to use a subspace V_n of H with $\dim V_n = n$ and to project the problem onto V_n, i.e. to solve the linear system

$$\langle A_\eta x_n, v_n\rangle = \langle y_\delta, v_n\rangle, \quad \forall v_n \in V_n, \tag{5.1.1}$$

for $x_n \in V_n$. Note that it suffices to test the equation only with basis vectors $v_n = e_i$, $i = 1, \dots, n$, of V_n. If A_η is positive definite on H, it is also positive definite on V_n, hence the condition (5.1.1) reduces to a uniquely solvable $n \times n$-system of linear equations. If $P_n : H \to V_n$ denotes the orthogonal projection onto V_n, then we may equivalently express the solution by

$$x_n = (P_n A_\eta|_{V_n})^{-1} P_n y_\delta. \tag{5.1.2}$$

Intuitively, a scale of spaces $(V_n)_{n\in\mathbb{N}}$ should be used that exhausts H and approximates any vector in H the better the larger n is. In the above paradigm we would set $h = n^{-1}$ because for $n \to \infty$ the approximation error decreases, while the operator norm $\|(P_n A_\eta|_{V_n})^{-1}\|$ increases.

The whole analysis of the error $x_n - x$ will be based upon the next theorem, which has inherited many ideas from an analogous statement by Hämarik (1983).

Theorem 3. *Let A be a selfadjoint and strictly positive definite operator on the Hilbert space H and let A_η be another linear operator on H with $\|A - A_\eta\| \le \eta$. Let $x, y, y_\delta \in H$ be given with $Ax = y$. Denote the orthogonal projection on a finite-dimensional subspace $V_n \subset H$ by P_n and introduce the operator $R_n := (P_n A|_{V_n})^{-1} P_n$.*

If $\eta < \|R_n\|^{-1}$ holds, then $P_n A_\eta$ is invertible on V_n. In this case, setting $R_{n\eta} := (P_n A_\eta|_{V_n})^{-1} P_n$, we obtain the estimate

$$\|R_{n\eta}\| \le \frac{\|R_n\|}{1 - \eta\|R_n\|}. \tag{5.1.3}$$

Furthermore, if $P_n A_\eta$ is invertible on V_n, then the Galerkin solution $x_n := R_{n\eta} y_\delta$ satisfies the error bound

$$\begin{aligned} \|x_n - x\| \le{}& \big[1 + \|R_{n\eta}\|(\|(\mathrm{Id} - P_n)A\| + \eta)\big]\|(\mathrm{Id} - P_n)x\| \\ &+ (1 + \|R_{n\eta}\|\eta)\|R_n(A_\eta x - y_\delta)\|. \end{aligned} \tag{5.1.4}$$

Proof. Due to $\langle P_n A v_n, v_n\rangle = \langle A v_n, v_n\rangle > 0$ for $v_n \in V_n \setminus \{0\}$, $P_n A$ is injective, hence bijective on V_n and $\|(P_n A|_{V_n})^{-1}\| = \sup_{\|v_n\|=1} \|P_n A v_n\|^{-1}$ follows. Note that the orthogonal projection P_n has always operator norm 1 and its norm is attained on V_n so that $\|R_n\|_{L^2 \to V_n} = \|(P_n A|_{V_n})^{-1}\|_{V_n \to V_n}$ holds. Therefore the condition $\eta < \|R_n\|^{-1}$ implies for $v_n \in V_n$, $\|v_n\| = 1$,

$$\|P_n A_\eta v_n\| \ge \|P_n A v_n\| - \|P_n(A_\eta - A)v_n\| \ge (\|(P_n A|_{V_n})^{-1}\|^{-1} - \eta) > 0.$$

Hence $P_n A_\eta|_{V_n}$ is injective and $R_{n\eta}$ exists. The bijectivity of $P_n A_\eta|_{V_n}$ yields the announced bound:

$$\|R_{n\eta}\| \le \sup_{\|v_n\|=1} \|P_n A_\eta v_n\|^{-1}\|P_n\| \le (\|R_n\|^{-1} - \eta)^{-1} = \frac{\|R_n\|}{1 - \eta\|R_n\|}.$$

The main estimate (5.1.4) is obtained by several applications of the triangle inequality. As additional ingredients in the proof we use that $R_{n\eta} A_\eta$ is a projection on V_n (the so called Galerkin projection), that the projection $\mathrm{Id} - P_n$ equals $(\mathrm{Id} - P_n)^2$

and that the selfadjointness of A and P_n implies $\|A(\mathrm{Id}-P_n)\| = \|(\mathrm{Id}-P_n)A\|$. We arrive at

$$\begin{aligned}
\|x - x_n\| &\leq \|x - P_n x\| + \|P_n x - R_{n\eta} A_\eta x\| + \|R_{n\eta}(A_\eta x - y_\delta)\| \\
&\leq \|(\mathrm{Id}-P_n)x\| + \|R_{n\eta} A_\eta (P_n - \mathrm{Id})x\| \\
&\quad + \|(R_{n\eta} - R_n)(A_\eta x - y_\delta)\| + \|R_n(A_\eta x - y_\delta)\| \\
&\leq \big[1 + \|R_{n\eta}\|(\|A(\mathrm{Id}-P_n)\| + \|(A_\eta - A)(\mathrm{Id}-P_n)\|)\big]\|(\mathrm{Id}-P_n)x\| \\
&\quad + \|R_{n\eta}(A - A_\eta)R_n(A_\eta x - y_\delta)\| + \|R_n(A_\eta x - y_\delta)\| \\
&\leq \big[1 + \|R_{n\eta}\|(\|(\mathrm{Id}-P_n)A\| + \eta)\big]\|(\mathrm{Id}-P_n)x\| \\
&\quad + (1 + \|R_{n\eta}\|\eta)\|R_n(A_\eta x - y_\delta)\|.
\end{aligned}$$

□

5.2 Construction of the estimator

A reason why we do not try to estimate arbitrary weights $a \in M([-r,0])$ is that the space $M([-r,0])$ with the total variation norm is not separable. Therefore any union $\bigcup_n V_n$ of finite-dimensional subspaces V_n cannot be dense in the norm topology. A Galerkin-type estimator can thus only work for a separable subspace of $M([-r,0])$, which is known to contain the weight. By assuming that the true underlying M^--weight a has a square-integrable Lebesgue density, we fix exactly such a separable subset. Moreover, we take advantage of the Hilbert space structure. It should already be mentioned that in Section 5.5 the properties of the estimator under construction are studied with respect to weak convergence if any measure $a \in M^-$ is the true underlying weight.

Assumption 1. *We assume that the weight $a \in M^-([-r,0])$ in the SDDE (2.2.10) has a Lebesgue density $g \in L^2([-r,0])$. The corresponding quantities will be indexed by g rather than by a.*

A scale of good approximation spaces (V_n) for $L^2([-r,0])$, as for instance splines with n uniformly spaced knots, is easily available (cf. the forthcoming Definition 7). We define the linear Galerkin estimator $\hat{g}_{T,n} \in V_n$ of g, measurable with respect to $\sigma(X(t),\ -r \leq t \leq T)$, as the solution of

$$\langle Q_T \hat{g}_{T,n}, f_n\rangle = \langle b_T, f_n\rangle, \quad \forall f_n \in V_n. \tag{5.2.5}$$

As remarked earlier, this can be rewritten more compactly as

$$\hat{g}_{T,n} = (P_n Q_T|_{V_n})^{-1} P_n b_T$$

with the orthogonal projection $P_n : L^2([-r,0]) \to V_n$. First establish the relationship with the abstract method developed in the previous section. Our Hilbert space is $L^2([-r,0])$, the true operator A is given by Q_g, the true data by $Q_g g$ and the perturbed quantities are $A_\eta = \frac{1}{T}Q_T$ and $y_\delta = \frac{1}{T}b_T$. Then condition (5.2.5) equals (5.1.1) after multiplication with T.

It was tacitly assumed in (5.2.5) that $P_n Q_T$ is invertible on V_n. We shall see in the proof of Theorem 4 that the condition in Theorem 3 for invertibility will be satisfied with a probability tending to one as $T \to \infty$, if the true weight function g is in $H^s([-r,0])$ for some $s > \frac{1}{2}$. We shall now, however, show directly that $P_n Q_T$ is almost surely invertible on any finite dimensional subspace V_n of $M([-r,0])$ and for any fixed $T > 0$. First an almost obvious lemma from linear algebra is needed.

Lemma 9. *Let n linear independent functions $f_j : [A,B] \to \mathbb{R}$, $j = 1,\dots,n$, $B > A$, be given. Define for $t \geq 0$ the shift operator to the right T_t by*

$$T_t f(s) = \begin{cases} 0, & \text{if } s \in [A, A+t) \\ f(s-t), & \text{if } s \in [A+t, B] \end{cases}, \quad s \in [A,B].$$

Suppose that there exists an $\varepsilon > 0$ with $f_j|_{[B-\varepsilon,B]} = 0$ for all $j = 1,\dots,n$. Then for any $m \in \mathbb{N}$ there are points $0 = t_1 < t_2 < \ldots < t_m < B - A$ such that the family of functions $(T_{t_i} f_j)_{\substack{1\leq i \leq m \\ 1 \leq j \leq n}}$ is linearly independent on $[A,B]$.

Proof. It suffices to show the linear independence of the family $(T_{t_i} f_j)$ for $m = 2$ with $t_2 =: \tau < \varepsilon$ since then the family $F = \{f_1,\dots,f_n, T_\tau f_1, \dots, T_\tau f_n\}$ satisfies the hypothesis of the lemma with $\varepsilon - \tau$ instead of ε and a simple induction yields the general result for $m > 2$.

For $t \in (A,B]$ consider the linear (!) subspace

$$N_t := \Big\{ \lambda = (\lambda_1,\dots,\lambda_n) \in \mathbb{R}^n \,\Big|\, \sum_{j=1}^n \lambda_j f_j|_{[A,t]} = \mathbf{0} \Big\} \subset \mathbb{R}^n \,.$$

Note the monotonicity property $N_t \subset N_s$ for $t > s$ and the identity $N_{B-\varepsilon} = \{0\}$ (by linear independence). We introduce the points τ_k where the dimension of N_t drops, i.e. $\tau_k := \sup\{t \in [A,B] \,|\, \dim N_t \geq k\}$, $k = 0,\dots,n$. We shall prove now that any choice of τ in the set

$$\{x \in (0,\varepsilon) \,|\, \forall\, 0 \leq k < l \leq n : \; x \neq \tau_k - \tau_l\}$$

will produce a linearly independent family F. To this end note that F is linearly independent if and only if the condition

$$\forall\, g = \sum_{j=1}^n \gamma_j f_j, \; \forall\, h = \sum_{j=1} \eta_j T_\tau f_j : \; g + h = 0 \Rightarrow g = h = 0$$

is satisfied, since $\{f_1,\dots,f_n\}$ and $\{T_\tau f_1,\dots,T_\tau f_n\}$ are each linearly independent families for $\tau < \varepsilon$.

Assume on the contrary that $g+h = 0$, but $g \neq 0$ and $h \neq 0$ and set $\rho := \inf\{s \in [A,B] \,|\, h(s) \neq 0\}$. Then $\rho \geq \tau$ holds by definition of h and $g + h = 0$ implies that g vanishes on the interval $[0,\rho]$. Therefore the coefficient vector $\gamma = (\gamma_1,\dots,\gamma_n)$ of g is an element of N_ρ. Thus, γ even lies in $N_{\tau_{k_0}}$ for some $\tau_{k_0} \geq \rho$. From this and $g + h = 0$ we conclude that h also vanishes on the interval $[A, \tau_{k_0}]$, which shows $\tau_{k_0} = \rho$ and $\eta \in N_{\rho-\tau}$. This is impossible because then η would lie in some $N_{\tau_{k_1}}$, $\tau_{k_1} \geq \rho - \tau$, but $\tau_{k_1} \neq \rho - \tau$ by the choice of τ so that h would vanish on the interval $[A, \tau_{k_1} + \tau]$, which is strictly larger than $[A,\rho]$. This contradiction shows that the family F is indeed linearly independent. □

Proposition 8. *Let $V_n \subset M([-r,0])$ be an n-dimensional subspace and $T > 0$. Then Q_T induces a strictly positive definite bilinear form on V_n with probability one; i.e. for any M^--weight a holds*

$$\mathbb{P}_a(\forall \mu_n \in V_n \setminus \{0\} : \; \langle Q_T \mu_n, \mu_n \rangle > 0) = 1.$$

Proof. Let the points $0 < t_1 < \ldots < t_n < T$ be fixed and be chosen later. Then, by the positive semidefiniteness of the form (cf. the second line of the calculations that follow) the following probability has to be shown to vanish:

$$\begin{aligned}
&\mathbb{P}_a(\exists \mu_n \in V_n \setminus \{0\} : \langle Q_T \mu_n, \mu_n \rangle = 0) \\
&= \mathbb{P}_a \left(\exists \mu_n \in V_n \setminus \{0\} : \int_0^T \left(\int_{-r}^0 X(t+u)\, d\mu_n(u) \right)^2 dt = 0 \right) \\
&= \mathbb{P}_a \left(\exists \mu_n \in V_n \setminus \{0\}\, \forall t \in [0,T] : \int_{-r}^0 X(t+u)\, d\mu_n(u) = 0 \right) \\
&\le \mathbb{P}_a \left(\exists \mu_n \in V_n \setminus \{0\}\, \forall i = 1,\dots,n : \int_{-r}^0 X(t_i+u)\, d\mu_n(u) = 0 \right) \\
&= \mathbb{P}_a \left(\text{the matrix } \mathcal{M} := \left(\int_{-r}^0 X(t_i+u)\, de_j(u) \right)_{1 \le i,j \le n} \text{ is singular} \right),
\end{aligned}$$

where $(e_1,\dots,e_n)$ is a basis of V_n and in the third line the a.s.-continuity of X was used. On $C([-r+t_1,T])$ the law $\mathbb{P}_g$ is equivalent to $\mathbb{P}_B$, the law of Brownian motion $(B(t),\, -r \le t \le T)$ starting at $t=-r$ in zero (cf. Theorem 2). It thus suffices to show that the matrix $\mathcal{M}$ is $\mathbb{P}_B$-almost surely non-singular. This will be accomplished by showing that the distribution of $\mathcal{M}$ is absolutely continuous with respect to the Lebesgue measure in $\mathbb{R}^{n\times n}$, since the set of non-singular matrices is an open and dense subset of $\mathbb{R}^{n\times n}$ and thus the singular matrices form a Lebesgue null set.

The problem is therefore reduced to showing that the covariance matrix $C \in \mathbb{R}^{n^2\times n^2}$ of the Gaussian vector $\mathcal{M} \in \mathbb{R}^{n^2}$ is non-singular. Suppose on the contrary that C is singular. Then there is a nonzero vector $(\alpha_{ij})_{1\le i,j\le n}$ such that

$$\begin{aligned}
0 &= \langle C(\alpha_{ij}), (\alpha_{ij}) \rangle_{\mathbb{R}^{n^2}} \\
&= \mathbb{E}_B[\langle (M_{ij}), (\alpha_{ij}) \rangle^2_{\mathbb{R}^{n^2}}] \\
&= \sum_{i,j,k,l=1}^n \alpha_{ij}\alpha_{kl} \int_{-r}^0 \int_{-r}^0 (r + \min(t_i+u, t_k+v))\, de_j(u)\, de_l(v) \\
&= \sum_{i,j,k,l=1}^n \alpha_{ij}\alpha_{kl} \int_{-r}^0 \int_{-r}^0 \int_{-r}^T \mathbf{1}_{[-r,t_i+u]}(s) \mathbf{1}_{[-r,t_k+v]}(s)\, ds\, de_j(u)\, de_l(v) \\
&= \int_{-r}^T \left(\sum_{i,j=1}^n \alpha_{ij} \int_{-r}^0 \mathbf{1}_{[-r,t_i+u]}(s)\, de_j(u) \right)^2 ds \\
&= \int_{-r}^T \left(\sum_{i,j=1}^n \alpha_{ij} e_j([s-t_i, 0]) \right)^2 ds,
\end{aligned}$$

where $[s-t_i,0]$ denotes the empty set for $s > t_i$ and $e_j(A) := e_j(A \cap [-r,0])$ for general real Borel sets A. Then the functions $f_j(u) := e_j([u,0])$ have just been shown to form a linearly dependent family $(f_j(\bullet - t_i))_{ij}$ of functions on $[-r,T]$ (use the left-continuity of f_j to get rid of the Lebesgue null set). On the other hand, the measures (e_j) are linearly independent in $M([-r,0])$ and hence so are their distribution functions f_j on $[-r,0]$. Therefore, an application of the preceding Lemma 9 with $A=-r$, $B=T$ and $\varepsilon=T$ yields the desired contradiction whenever the points $t_1,\dots,t_n$ are chosen according to this lemma. Note that this choice only depends on the basis (e_j) and not on X. □

Remarks 7.

- *This proposition is an improvement of the result in Küchler and Sørensen (1997, Lemma 9.1.2), which is stated for point measures μ_n and effectively only shows that any nonzero vector almost surely never lies in the kernel of Q_T so that the null set involved might still be vector-dependent, hence not universal.*

- *The general question, whether Q_T as an operator on $M([-r,0])$ or just on $L^2([-r,0])$ is almost surely injective, remains a very interesting open problem. By the Girsanov theorem, it can be reformulated: Given a standard Brownian motion $(B(t),\, t \in [0,T])$, $T > 1$, what is the probability that for $\tau \in [0, T-1]$ the segments $B_\tau = (B(s+\tau), s \in [0,1])$ span a dense subspace of $C([0,1])$ or $L^2([0,1])$ respectively? One can prove that because of the independence of increments the segments B_τ for $\tau \in \mathbb{N}$ span a dense subspace in $C([0,1])$ with probability one. This yields the answer in the limiting case $T = \infty$. The finite-dimensional proof relies heavily on the fact that the set of singular matrices has Lebesgue measure zero so that an infinite-dimensional adaptation does not seem feasible.*

5.3 Upper bounds for the L^2-risk

In this section we shall assess the quality of an estimator $\hat{g}$ of the weight function g by the $L^2(\Omega \times [-r,0])$-risk $\mathbb{E}_g[\|\hat{g} - g\|^2_{L^2[-r,0]}]^{1/2}$. The use of the $L^2([-r,0])$-loss function is quite natural since the Galerkin method essentially relies on a Hilbert space structure. Considering only second moments with respect to the probability measure allows for straightforward calculations since the error term $\frac{1}{\sqrt{T}}(b_T - Q_T a)$ has the same covariance structure as the Gaussian process X. The estimation of higher moments is technically much more involved, as will be seen in the chapter on testing theory where fourth order moments need to be estimated.

We have to choose the scale of approximation spaces $(V_n)_{n\in\mathbb{N}}$ carefully. As is well known from approximation theory, the so-called Bernstein and Jackson inequalities or direct and inverse estimates are fundamental. The first one ensures a certain order of approximation as $n \to \infty$, the second one gives a bound on the decay of the smallest eigenvalue of the covariance operator restricted to V_n. Expressed in statistical terminology, the Jackson inequality bounds the bias, whereas the Bernstein inequality helps to control the variance term.

Definition 7. *A sequence $(V_n)_{n\in\mathbb{N}}$ of n-dimensional subspaces of $L^2([-r,0])$ will be called s-*approximating *if it satisfies the following inequalities with fixed constants $C_J, C_B > 0$, P_n denoting the orthogonal projection onto V_n:*

$$\|(\mathrm{Id} - P_n)u\|_{L^2} \le C_J \|u\|_{H^\alpha} n^{-\alpha}, \quad \forall u \in H^\alpha([-r,0]),\ \alpha \in \{s, 2\}, \tag{5.3.6}$$

$$\langle Q v_n, v_n \rangle \ge C_B n^{-2} \|v_n\|^2_{L^2}, \quad \forall v_n \in V_n. \tag{5.3.7}$$

Here, Q denotes any covariance operator Q_a for $a \in M^-([-r,0])$ or the operator Q_W from Remark 5.

Remark 7. *It follows from Proposition 4 and Remark 5 that (5.3.7) holds for all $(Q_a)_{a\in M^-}$, possibly with a different constant, once it is established for one Q_a or for Q_W.*

The Jackson inequality (5.3.6) is satisfied by any reasonable and smooth enough finite element space or wavelet multiresolution analysis. If the operator Q were an isomorphism from $L^2([-r,0])$ to $H^2([-r,0])$ the Bernstein-type inequality (5.3.7) would also be straight forward. Since Q is not surjective, we have to work a little bit harder, but it turns out that the inequality can also be established under usual conditions by some refined duality argument.

Proposition 9. *If the scale of subspaces $(V_n)_{n\in\mathbb{N}}$ is contained in $H^1([-r,0])$ and satisfies*

$$\lim_{n\to\infty}\|f-P_nf\|_{H^1}=0,\quad \forall\, f\in H^1([-r,0]),$$
$$\|v_n\|_{H^1}\lesssim n\|v_n\|_{L^2},\quad \forall\, v_n\in V_n,$$

then the Bernstein inequality (5.3.7) is satisfied by $(V_n)_{n\in\mathbb{N}}$.

Proof. We use Q_W to prove (5.3.7) because its square root $Q_W^{1/2}$ can be determined explicitly. Since Q_W is a positive definite compact operator on $L^2([-r,0])$, there exists a decreasing sequence (λ_n) of positive eigenvalues and a corresponding orthonormal basis of eigenfunctions (e_n) such that the spectral decomposition

$$Q_Wf=\sum_{n=1}^{\infty}\lambda_n\langle f,e_n\rangle e_n,\quad f\in L^2([-r,0]),$$

is valid Dunford and Schwartz (1963, Cor. X.3.5). The eigenfunctions must satisfy the differential equation (D denotes the derivative operator)

$$\lambda_nD^2e_n=D^2Q_We_n=D^2\left(\int_{-r}^{0}(\min(s,\bullet)+r+1)e_n(s)\,ds\right)=-e_n.$$

Therefore e_n must be a trigonometric function of the form

$$e_n(s)=A_n\cos(\omega_ns)+B_n\sin(\omega_ns)$$

with $\omega_n=\lambda_n^{-1/2}$. Using $(DQ_We_n)=\int_\bullet^0 e_n(s)ds$ we obtain the restriction $e_n'(0)=0$ and thus $B_n=0$. An integration by parts shows

$$Q_W(\cos(\omega_n\bullet))(t)=\omega_n^{-2}(\cos(\omega_nt)+(2r+1)\omega_n\sin(\omega_nr)-\cos(\omega_nr)).$$

For e_n being an eigenfunction it is therefore necessary and sufficient that ω_n satisfies $(2r+1)\omega_n\sin(\omega_nr)=\cos(\omega_nr)$, which by the periodicity of the tangens yields $\omega_n\sim\frac{\pi}{r}n$ and hence $\lambda_n\sim n^{-2}$. The exponent 2, by the way, corresponds to the degree of illposedness of the operator Q_W. Finally set $A_n=\|\cos(\omega_n\bullet)\|_{L^2}^{-1}$.

Using the spectral decomposition, we have the following formal representation of $DQ_W^{1/2}$

$$DQ_W^{1/2}f=D\left(\sum_{n=1}^{\infty}\lambda_n^{1/2}\langle f,e_n\rangle e_n\right)=-\sum_{n=1}^{\infty}\langle f,e_n\rangle A_n\sin(\omega_n\bullet),\quad f\in L^2([-r,0]).$$

In order to justify this formal expansion, we first calculate the L^2-scalar product $\langle\sin(\omega_m\bullet),\sin(\omega_n\bullet)\rangle$ for $m\neq n$:

$$\begin{aligned}\int_{-r}^{0}\sin(\omega_mx)\sin(\omega_nx)\,dx&=\frac{-1}{\omega_n}\sin(\omega_mr)\cos(\omega_nr)+\frac{\omega_m}{\omega_n}\int_{-r}^{0}\cos(\omega_mx)\cos(\omega_nx)\,dx\\&=-(2r+1)\sin(\omega_mr)\sin(\omega_nr)+0.\end{aligned}$$

Equally, we obtain $\|\sin(\omega_n\bullet)\|_{L^2}^2=-(2r+1)\sin^2(\omega_nr)+A_n^{-2}$. Due to $\langle\mathbf{1},e_n\rangle=A_n\omega_n^{-1}\sin(\omega_nr)$ for all $n\in\mathbb{N}$ we find for $f\in L^2([-r,0])$

$$\begin{aligned}\|DQ_W^{1/2}f\|_{L^2}^2&=-\sum_{m,n=1}^{\infty}\langle f,e_m\rangle\langle f,e_n\rangle A_mA_n(2r+1)\sin(\omega_mr)\sin(\omega_nr)+\sum_{n=1}^{\infty}\langle f,e_n\rangle^2\\&=\sum_{n=1}^{\infty}\langle f,e_n\rangle^2-(2r+1)\left(\sum_{n=1}^{\infty}A_n\sin(\omega_nr)\langle f,e_n\rangle\right)^2\\&=\|f\|_{L^2}^2-(2r+1)\langle f,Q_W^{1/2}\mathbf{1}\rangle^2,\end{aligned}$$

which a posteriori justifies the expansion of $DQ_W^{1/2}$ in an L^2-sense. Let V be the space that is L^2-orthogonal to $Q_W^{1/2}\mathbf{1}$. Then any function $f \in V$ satisfies $\|Q_W^{1/2} f\|_{H^1} \geq \|f\|_{L^2}$ and $Q_W^{1/2}(V)$ is closed in H^1. Hence, the whole range $Q_W^{1/2}(L^2)$ is the sum of a closed subspace and a one-dimensional subspace of H^1, which is always closed Rudin (1991, Thm. 1.42). By the open mapping theorem $Q_W^{1/2} : L^2 \to Q_W^{1/2}(L^2)$ is an isomorphism and there is a constant $c > 0$ such that $\|Q_W^{1/2} f\|_{H^1} \geq c\|f\|_{L^2}$ holds for all $f \in L^2([-r, 0])$ Rudin (1991, Cor. 2.12c).

For the final estimate we shall additionally use the fact that the range of $Q_W^{1/2}$ is dense in $L^2([-r, 0])$ due to $e_n \in Q_W^{1/2}(L^2)$ for all $n \in \mathbb{N}$. Hence P_n maps $Q_W^{1/2}(L^2)$ onto V_n. By a duality argument we obtain uniformly in n, using the two assumptions on (V_n) in the statement of the proposition,

$$
\begin{aligned}
\inf_{v_n \in V_n \setminus \{0\}} \frac{\langle Q_W v_n, v_n\rangle}{\|v_n\|_{L^2}^2} &= \inf_{v_n \in V_n \setminus \{0\}} \frac{\|Q_W^{1/2} v_n\|_{L^2}^2}{\|v_n\|_{L^2}^2} \\
&= \inf_{v_n \in V_n \setminus \{0\}} \sup_{f \in L^2 \setminus \{0\}} \frac{\langle Q_W^{1/2} v_n, f\rangle^2}{\|f\|_{L^2}^2 \|v_n\|_{L^2}^2} \\
&= \inf_{v_n \in V_n \setminus \{0\}} \sup_{f \in L^2 \setminus \{0\}} \frac{\langle v_n, Q_W^{1/2} f\rangle^2}{\|f\|_{L^2}^2 \|v_n\|_{L^2}^2} \\
&= \inf_{v_n \in V_n \setminus \{0\}} \sup_{h \in Q_W^{1/2}(L^2) \setminus \{0\}} \frac{\langle v_n, h\rangle^2}{\|Q_W^{-1/2} h\|_{L^2}^2 \|v_n\|_{L^2}^2} \\
&\geq \inf_{v_n \in V_n \setminus \{0\}} \sup_{h \in Q_W^{1/2}(L^2) \setminus \{0\}} \frac{\langle v_n, P_n h\rangle^2}{c^{-2}\|h\|_{H^1}^2 \|v_n\|_{L^2}^2} \\
&\gtrsim \inf_{v_n \in V_n \setminus \{0\}} \sup_{\substack{h \in Q_W^{1/2}(L^2) \\ P_n h \neq 0}} \frac{\langle v_n, P_n h\rangle^2}{\|P_n h\|_{H^1}^2 \|v_n\|_{L^2}^2} \\
&= \inf_{v_n \in V_n \setminus \{0\}} \sup_{h_n \in V_n \setminus \{0\}} \frac{\langle v_n, h_n\rangle^2}{\|h_n\|_{H^1}^2 \|v_n\|_{L^2}^2} \\
&\geq \inf_{v_n \in V_n \setminus \{0\}} \frac{\|v_n\|_{L^2}^2}{\|v_n\|_{H^1}^2} \\
&\gtrsim n^{-2}.
\end{aligned}
$$

□

Example 6. *A whole class of s-approximating sequences of approximation spaces is provided by splines of order $m \geq s \vee 2$ with equidistant knots Schumaker (1981).*

Wavelet multiresolution analyses (V_j) in $L^2([-r, 0])$ are also s-approximating whenever they are $(s \vee 2)$-regular (Corollary 12).

We introduce the class of weight functions considered and prove an asymptotic risk upper bound uniformly for weights from this class.

Definition 8. *For $s > 0$, $S > 0$ and $\delta > 0$ set*

$$M(s, S, \delta) := \{g \in H^s([-r, 0]) \mid \|g\|_s \leq S,\ v_0(g) \leq -\delta\}.$$

Proposition 10. *Assume that the subspaces (V_n) are s-approximating and that $\hat{g}_{T,n}$ is determined by (5.2.5). Then $\hat{g}_{T,n}$ is $\sigma(X(t), -r \leq t \leq T)$-measurable. Introduce the random set*

$$\mathcal{B}_g := \{\|Q_g - \tfrac{1}{T} Q_T\| \|(P_n Q_g|_{V_n})^{-1}\| \leq \tfrac{1}{2}\}.$$

If $g \in H^s([-r,0])$ with $v_0(g) < 0$ is the true underlying weight function of the stationary solution of the SDDE (2.2.10), then we obtain the bound

$$\mathbb{E}_g[\|\hat{g}_{T,n} - g\|_{L^2}^2 \mathbf{1}_{\mathcal{B}_g}] \lesssim n^{-2s} + n^3 T^{-1}. \tag{5.3.8}$$

The constant may be chosen uniformly for $g \in M(s,S,\delta)$ for fixed $s > 0$, $S > 0$ and $\delta > 0$.

Proof. The quantity $\hat{g}_{T,n} \in V_n \subset L^2([-r,0])$ is a continuous function of Q_T and b_T with respect to the operator norm and the L^2-norm, respectively. Since $q_T \mapsto Q_T$ is continuous with respect to the L^2-norm and the operator norm and since q_T and b_T are $\sigma(X(t), -r \leq t \leq T)$-measurable, so is $\hat{g}_{T,n}$.

Note that the estimate (5.1.3) from Theorem 3 is satisfied on the set $\mathcal{B}_g$ (recall $A = Q_g$, $A_\eta = \frac{1}{T} Q_T$). Inserting this estimate into (5.1.4), we obtain on $\mathcal{B}_g$

$$\|\hat{g}_{T,n} - g\|_{L^2} \leq (2 + 2\|R_n\| \|(\mathrm{Id} - P_n) Q_g\|) \|(\mathrm{Id} - P_n) g\|_{L^2} + 2T^{-1} \|R_n(Q_T g - b_T)\|_{L^2}$$

with $R_n = (P_n Q_g|_{V_n})^{-1} P_n$. Because the subspaces V_n are s-approximating, the following inequalities hold:

$$\begin{aligned} \|(\mathrm{Id} - P_n) g\|_{L^2} &\leq C_J S n^{-s}, \\ \|(\mathrm{Id} - P_n) Q_g\| &\lesssim \|(\mathrm{Id} - P_n)\|_{H^2 \to L^2} \leq C_J n^{-2}, && (5.3.9) \\ \|R_n\| &= \inf_{\|v_n\|=1} \langle Q_g v_n, v_n \rangle \lesssim n^2. && (5.3.10) \end{aligned}$$

Note that in inequality (5.3.9) the constant depends weakly continuously on g by Theorem 5, hence may be chosen uniformly for all $g \in M(s,S,\delta)$ by the usual compactness argument. The constant from Bernstein's inequality in the last inequality (5.3.10) can also be chosen uniformly for all g with $\|g\|_{H^s} \leq S < \infty$ by Proposition 4 due to the weak compactness of H^s-balls. The estimate on $\mathcal{B}_g$ simplifies to

$$\|\hat{g}_{T,n} - g\|_{L^2} \lesssim (1 + n^2 n^{-2}) \|g\|_{H^s} n^{-s} + T^{-1} \|R_n(Q_T g - b_T)\|_{L^2}.$$

Only the second "variance" term is stochastic and needs further treatment. Denote by $(e_1, \ldots, e_n)$ an L^2-orthonormal basis of V_n. Then the selfadjointness of R_n, the Fubini theorem for stochastic integrals (use the continuity of X and Protter (1992, Thm. 46)) and inequality (5.3.10) give

$$\begin{aligned} \mathbb{E}_g[\|R_n(Q_T g - b_T)\|_{L^2}^2] &= \sum_{i=1}^n \mathbb{E}_g[\langle R_n(Q_T g - b_T), e_i \rangle^2] \\ &= \sum_{i=1}^n \mathbb{E}_g\left[\left(\int_0^T \langle R_n e_i, X(\bullet + t) \rangle \, dW(t)\right)^2\right] \\ &= \sum_{i=1}^n T\, \mathbb{E}_g[\langle R_n e_i, X(\bullet) \rangle^2] \\ &= T \sum_{i=1}^n \langle Q_g R_n e_i, R_n e_i \rangle \\ &= T \sum_{i=1}^n \langle P_n Q_g (P_n Q_g|_{V_n})^{-1} e_i, R_n e_i \rangle \\ &= T \sum_{i=1}^n \langle e_i, R_n e_i \rangle \\ &\lesssim T n^3. \end{aligned}$$

Finally, the general estimate $(A+B)^2 \le 2(A^2+B^2)$ yields the desired result

$$\mathbb{E}_g[\|\hat{g}_{T,n} - g\|\mathbf{1}^2_{\mathcal{B}_g}] \lesssim n^{-2s} + T^{-1}n^3$$

with a uniform constant for $g \in M(s, S, \delta)$. □

The preceding proposition states the classical result in nonparametric estimation theory that the error can be split into a bias term and a variance term. The normalised bias term is of order n^{-s} and comes from the fact that functions in H^s can be approximated in V_n with a rate n^{-s}. The variance term is classically of order $\mathbb{E}[\|P_n N\|^2_{L^2}]$ where N denotes the noise. In the abstract white noise model or the related density estimation or regression problems the noise is of the form εW, W white noise, and we obtain $\varepsilon^2 \dim(V_n)$ as variance term. In our case however, the noise $T^{-1}(Q_T g - b_T)$ resembles in a second order sense $T^{-1/2}X = T^{-1/2}Q_g^{1/2}W$ (cf. the above calculations), but due to the illposedness the application of Q_g^{-1} amplifies the noise to $T^{-1/2}Q_g^{-1/2}W$ and we obtain a variance term of order $T^{-1}n^3$. The factor n^3 can be understood as the dimension of V_n multiplied by n to the power of the degree of illposedness of Q_g or more succinctly as the trace norm of R_n. This is the intuitive explanation for the phenomenon that the variance term is of higher order than usual. For an abstract investigation of these questions we refer to Nussbaum and Pereverzev (1999).

Simulations confirm the rapid growth of the variance term for an increasing number of variables. Figure 5.3.1 shows the Galerkin estimator derived from the simulation of an SDDE for a period of length 2000 by a discretisation step $\Delta = 1/1600$. As initial condition an Ornstein-Uhlenbeck process has been chosen, which is close in law to the true stationary solution. For the Galerkin estimator only discrete observations of width $\Delta = 1/200$ were used in order to save computer capacity. The results show clearly that for a small number of knots the true exponential weight function cannot be well approximated (first row) and that for a large number of knots (last row) a lot of noise is present, which intuitively one would like to smooth out. In the second row there is already enough variability to well approximate the weight function, but still a sufficiently low dimension to keep the noise small. Starting with the largest number of knots, one can consider the estimators with less knots as an averaging of the high-dimensional estimator; so the smoothing takes place by a reduction of the approximation space used.

As usual, we balance the bias and the variance term by the right choice of n in order to obtain a convergence rate for $T \to \infty$. Observe, however, that we still have to get rid of the assumption concerning the random set $\mathcal{B}_g$. We shall show that for $g \in H^s([-r, 0])$ with $s > \frac{1}{2}$ the probability of $\mathcal{B}_g$ tends fast enough to one for $T \to \infty$, but on its complement we are not able to bound $\|R_{n\eta}\|$ and $\|\hat{g}_{T,n} - g\|$ may explode. This is why we artificially renormalise $\hat{g}_{T,n}$ in order to keep it bounded. We could equivalently measure the risk with the truncated risk function $\mathbb{E}_g[\min(\|\hat{g}_{T,n} - g\|, L)^2]^{1/2}$ for some $L > 0$.

Theorem 4. *Assume the hypotheses and definitions of the preceding Proposition 10, but rescale $\hat{g}_{T,n}$ to $S\|\hat{g}_{T,n}\|^{-1}_{L^2}\hat{g}_{T,n}$ in the case $\|\hat{g}_{T,n}\|_{L^2} > S$. Then the choice $n(T) \sim T^{\frac{1}{2s+3}}$ yields the uniform asymptotic upper bound $T^{-\frac{s}{2s+3}}$ for the risk. More precisely, for any $\delta > 0$, any $s > \frac{1}{2}$ and any $S > 0$ the following holds:*

$$\sup_{g \in M(s,S,\delta)} \mathbb{E}_g[\|\hat{g}_{T,n(T)} - g\|^2_{L^2}]^{1/2} \lesssim T^{-\frac{s}{2s+3}}.$$

Proof. Observe first that the rescaling always produces a better estimate of g, since we know a priori $\|g\|_{L^2} \le \|g\|_s \le S$. By the choice of $n(T)$ the estimate 5.3.8 may

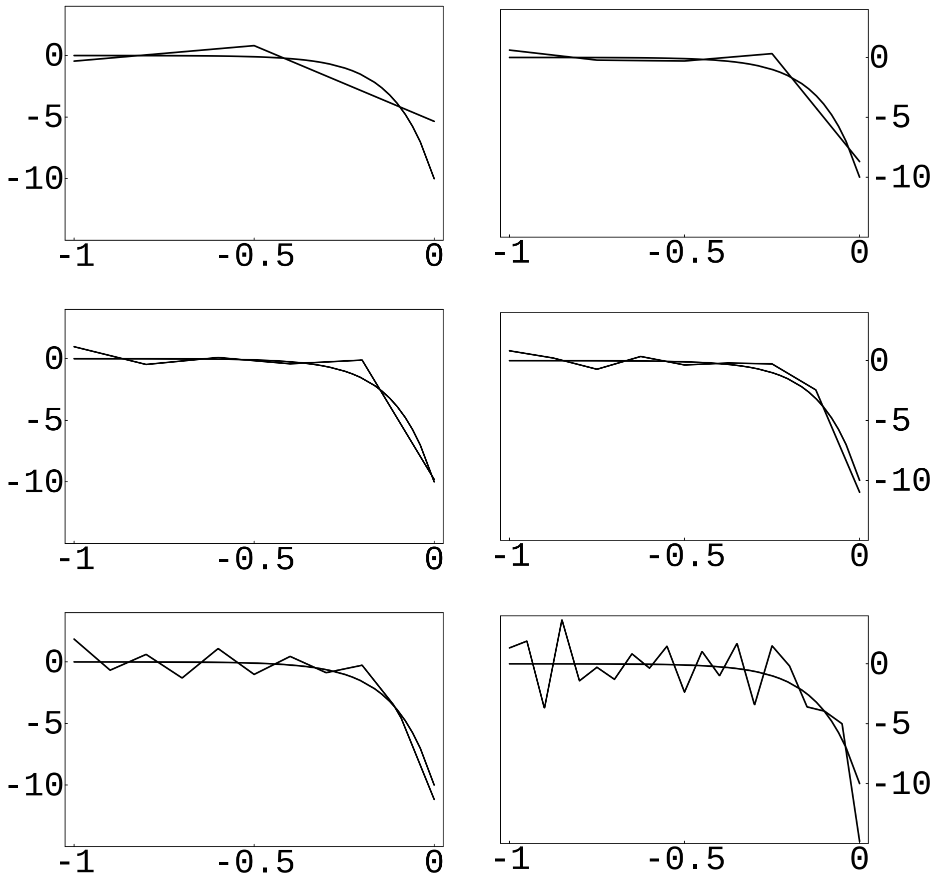

Figure 5.3.1: The weight function $g(t) = -10^{1+3t}$, $t \in [-1, 0]$, and some simulated Galerkin estimators. In the different rows linear splines with 3 and 5, 6 and 9, 11 and 21 uniform knots were used.

be rewritten as the uniform estimate

$$\sup_{g \in M(s,S,\delta)} \mathbb{E}_g[\|\hat{g}_{T,n(T)} - g\|_{L^2}^2 \mathbf{1}_{\mathcal{B}_g}]^{1/2} \lesssim T^{-\frac{s}{2s+3}}.$$

Due to the rescaling, $\|\hat{g}_{T,n}\|_{L^2} \le S$ is guaranteed, so that it suffices to prove $\mathbb{P}_g(\Omega \setminus \mathcal{B}_g) \lesssim T^{-\frac{2s}{2s+3}}$ uniformly in g to finish the proof. From Proposition 5 we infer $\mathbb{E}_g[\|q_g - \frac{1}{T} q_T\|_{L^2}^{2m}] \lesssim T^{-m}$ uniformly for g in $M(s, S, \delta) \subset M(S, \delta)$. Since we have established the estimate $\|R_{n(T)}\|^{2m} \lesssim n(T)^{4m} \sim T^{\frac{4m}{2s+3}}$ in inequality (5.3.10) uniformly for $g \in M(s, S, \delta)$, we obtain

$$\limsup_{T\to\infty} T^{\frac{(2s+3)m-4m}{2s+3}} \sup_{g \in M(s,S,\delta)} \mathbb{E}_g[\|q_g - \tfrac{1}{T} q_T\|^{2m} \|R_{n(T)}\|^{2m}] < \infty.$$

Thus, by the generalized Chebyshev inequality and by the general norm bound $\|Q_g - \frac{1}{T} Q_T\| \le \|q_g - \frac{1}{T} q_T\|_{L^2}$ we find

$$\limsup_{T\to\infty} \Big(T^{\frac{2sm-m}{2s+3}} \sup_{g \in M(s,S,\delta)} \mathbb{P}_g(\Omega \setminus \mathcal{B}_g)\Big) < \infty.$$

The choice of m is at our disposal and the last estimate proves that for $s > \frac{1}{2}$ the probability $\mathbb{P}_g(\Omega \setminus \mathcal{B}_g)$ tends to zero for $T \to \infty$ faster than any polynomial in T and uniformly in g. □

Remarks 8.

- *A look at Corollary 2 in conjunction with Remark 2 shows that the Galerkin estimator $\hat{g}_{T,n}$ is asymptotically the maximum likelihood estimator of the weight g, if g is supposed to lie in the finite-dimensional space V_n. The proposed nonparametric estimator $\hat{g}_{T,n(T)}$ may hence be interpreted as a maximum likelihood estimator for a misspecified parametric model where the misspecification (the bias) shrinks with increasing observation time. This relationship with the maximum likelihood estimator already indicates why the Galerkin estimator is rate-optimal in a minimax setting. Moreover, it gives the estimator $\hat{g}_{T,n}$ a statistical interpretation, even if we are not in the stationary regime, but still observe a trajectory of an affine SDDE, or if the model is misspecified in the sense of Section 5.5.*

- *Observe that a practical implementation of the Galerkin estimator is very easy. Only the quantities q_T and b_T have to be calculated, an approximation space V_n has to be chosen and the linear system (5.2.5) has to be inverted. The problems due to discrete observations are investigated in the Sections 5.4 and 6.2. Note that we have assumed to know the regularity s of the unknown weight function g in order to choose the right space V_n. This drawback will be removed by the adaptive method presented in Chapter 7. The constants involved are in practice difficult to determine so that the automatic calibration of bias and variance term might pose a problem. A more precise result, not only announcing the rate of convergence, but also the asymptotic constant (cf. Pinsker (1980) for a signal plus noise setting), could be strived for. However, this seems not to be easily achievable.*

- *The restriction $s > \frac{1}{2}$ was necessary in order to bound the probability of $\Omega \setminus \mathcal{B}_g$. For $s < \frac{1}{2}$ we still get convergence, but have to calibrate $n(T)$ differently and obtain a slower rate. The problem stems from the fact that we estimate $\|R_n\|_{L^2 \to L^2} \lesssim n^2$ deterministically, while – as seen in the estimate of the variance term – it should be more appropriate to use a mean square estimate of $\|R_n(Q_T - Q_g)\|_{L^2 \to L^2}$, which might be of order $n^{3/2}$ in n. It is thus likely that the upper bound holds for all $s > 0$ and the techniques employed are just too rough.*

- *At this point it is natural to ask how the bound on the risk changes with respect to a diffusion coefficient σ in (2.2.10), which is not equal to one. The – perhaps surprising – answer is that it is completely independent of $\sigma > 0$, since both Q_T and b_T would involve a factor of σ^2 such that it cancels out in the calculation of $\hat{g}_{T,n}$. This is in perfect agreement with the theory of estimation for autoregressive processes, where a discretized variant of the Galerkin estimator is given by the Yule-Walker estimator Brockwell and Davis (1996, Section 8.1). As was pointed out before, the diffusion coefficient merely scales the solution process and can be set to one by a linear space transformation.*

- *The functional law of the iterated logarithm for mixing sequences of Banach space-valued random variables Dehling and Philipp (1982) can be used to derive almost sure convergence results. For integer values of T the quantities q_T and b_T can be regarded as sums of functions of the stationary and β-mixing sequence $(X(n+s), 0 \le s \le 1)_{n \in \mathbb{N}}$ in $L^2([0,1])$. The techniques developed so far then readily yield an almost sure upper risk bound of $\hat{g}_{T,n}$ in $L^2([-r,0])$ of order $n^{-s} + T^{-1/2}n^4$ up to (iterated) logarithmic factors. It is very likely* *that the rate for the variance term can be decreased to $T^{-1/2}n^3$ times logarithmic factors. The main difficulty is a tight almost-sure estimate of the term $\|R_n(Q_T g - b_T)\|$.*

5.4 Discrete time observations ($\Delta \to 0$)

While the theoretical justification of the Euler scheme used for the generation of trajectories follows from Section 2.5, it remains to investigate how the Galerkin method applies to discrete data due to a numerical implementation or an observation at discrete time points only. Since the estimation method inherently deals with errors, it is no surprise that for discrete data at time points of maximal distance Δ the error does not increase significantly, if Δ is small.

We assume that a trajectory of an affine SDDE is observed at time points $0 = t_0 < t_1 < \cdots < t_N = T$ with maximal width

$$\Delta := \max_{1 \le i \le N} |t_i - t_{i-1}|. \tag{5.4.11}$$

Note that we have deliberately not chosen an equidistant time grid because continuous time models are particularly useful in the case of non-uniformly spaced data.

Definition 9. *For a stationary solution X of the affine SDDE (2.2.10) and a collection $\mathcal{T}^\Delta$ of time points with maximal width Δ as in (5.4.11) we put*

$$X^\Delta(t) := X(\lfloor t \rfloor_\Delta), \quad t \ge 0, \text{ where } \lfloor t \rfloor_\Delta := \max\{t_i \in \mathcal{T}^\Delta \mid t_i \le t\}.$$

The quantities q_T^Δ, Q_T^Δ and b_T^Δ are defined by

$$\begin{aligned} q_T^\Delta(u,v) &:= \int_0^T X^\Delta(t+u) X^\Delta(t+v)\, dt, \quad u, v \in [-r, 0], \\ Q_T^\Delta f(s) &:= \int_{-r}^0 q_T^\Delta(s,v) f(v)\, dv, \quad f \in L^1([-r,0]),\ s \in [-r,0], \\ b_T^\Delta(s) &:= \sum_{i=1}^N X^\Delta(t_{i-1}+s)(X(t_i) - X(t_{i-1})), \quad s \in [-r,0]. \end{aligned}$$

Lemma 10. *The following asymptotic upper bound holds for $m \in \mathbb{N}$:*

$$\mathbb{E}_a[\|q_T - q_T^\Delta\|^{2m}_{L^{2m}([-r,0]^2)}] \lesssim \Delta^{2m} T^{2m}.$$

Proof. We first split the error into a bias and a variance-type term:

$$\begin{aligned} &\mathbb{E}_a[\|q_T - q_T^\Delta\|^{2m}_{L^{2m}([-r,0]^2)}] \\ &= \int_{-r}^0 \int_{-r}^0 \mathbb{E}_a\left[\left(\int_0^T (X(t+u)X(t+v) - X^\Delta(t+u)X^\Delta(t+v))\, dt\right)^{2m}\right] du\, dv \\ &\lesssim \sup_{u,v} \left(\int_0^T (q_a(u-v) - q_a(\lfloor t+u \rfloor_\Delta - \lfloor t+v \rfloor_\Delta))\, dt\right)^{2m} \\ &\quad + \sup_{u,v} \mathbb{E}_a\Bigg[\Bigg(\int_0^T (X(t+u)X(t+v) - q_a(u-v))\, dt \\ &\qquad - \int_0^T (X^\Delta(t+u)X^\Delta(t+v) - q_a(\lfloor t+u \rfloor_\Delta - \lfloor t+v \rfloor_\Delta))\, dt\Bigg)^{2m}\Bigg]. \end{aligned}$$

The first term (the bias) is of order $(\Delta T)^{2m}$ since q_a is Lipschitz continuous. The second summand (the variance) resembles the moment estimate of Proposition 5. In fact, a completely analogous proof, which only relies on the regularity and decay property of the covariance function, shows that the second term is of order $T^m \Delta^{2m}$. In our case, however, a rough estimate using Jensen's inequality and Gaussian moment properties suffices to bound the second term by

$$\begin{aligned}
&\leq T^{2m} \sup_{u,v,t} \mathbb{E}_a \Big[\big(X(t+u)X(t+v) - q_a(u-v) - X^\Delta(t+u)X^\Delta(t+v) \\
&\qquad\qquad + q_a(\lfloor t+u \rfloor_\Delta - \lfloor t+v \rfloor_\Delta) \big)^{2m} \Big] \\
&\lesssim T^{2m} \sup_{u,v,t} (q_a(u-v) - q_a(\lfloor t+u \rfloor_\Delta - \lfloor t+v \rfloor_\Delta))^{2m} \\
&\lesssim T^{2m} \Delta^{2m}.
\end{aligned}$$

Thus, the asserted upper bound holds true. □

Lemma 11. *For $b_T^\Delta - b_T$ we obtain the asymptotic upper bound*

$$\mathbb{E}_a[\|b_T^\Delta - b_T\|^2_{L^2([-r,0])}] \lesssim \Delta T + \Delta^2 T^2.$$

Proof. We split again into a bias and a variance term. By the Lipschitz continuity of the covariance function we obtain as a bound for the bias term

$$\begin{aligned}
&\int_{-r}^0 (\mathbb{E}_a[b_T^\Delta(s) - b_T(s)])^2 \, ds \\
&\leq r \sup_s \left(\mathbb{E}_a \left[\sum_{i=1}^N \int_{t_{i-1}}^{t_i} (X(\lfloor t_{i-1}+s \rfloor_\Delta) - X(t+s)) \, dX(t) \right] \right)^2 \\
&= r \sup_s \left(\sum_{i=1}^N \int_{t_{i-1}}^{t_i} \int_{-r}^0 \mathbb{E}_a[(X(\lfloor t_{i-1}+s \rfloor_\Delta) - X(t+s))X(t+u)] \, da(u) \, dt \right)^2 \\
&\lesssim (\|a\|_{TV} T \Delta)^2.
\end{aligned}$$

The variance term may be treated, using separate estimates for the drift and the diffusion part:

$$\begin{aligned}
&\int_{-r}^0 \operatorname{Var}_a(b_T^\Delta(s) - b_T(s)) \, ds \\
&\leq r \sup_s \operatorname{Var}_a \left[\sum_{i=1}^N \int_{t_{i-1}}^{t_i} (X(\lfloor t_{i-1}+s \rfloor_\Delta) - X(t+s)) \, dX(t) \right] \\
&\lesssim \sup_{s,u_1,u_2} \sum_{i,j=1}^N \int_{t_{i-1}}^{t_i} \int_{t_{j-1}}^{t_j} \operatorname{Cov}_a \Big[(X(\lfloor t_{i-1}+s \rfloor_\Delta) - X(t^{(1)}+s))X(t^{(1)}+u_1), \\
&\qquad\qquad (X(\lfloor t_{j-1}+s \rfloor_\Delta) - X(t^{(2)}+s))X(t^{(2)}+u_2) \Big] \, dt^{(1)} \, dt^{(2)} \|a\|^2_{TV} \\
&\quad + \sup_s \sum_{i=1}^N \int_{t_{i-1}}^{t_i} \mathbb{E}_a[(X(\lfloor t_{i-1}+s \rfloor_\Delta) - X(t+s))^2] \, dt.
\end{aligned}$$

The diffusion part (last summand) is of order ΔT due to the Lipschitz continuity of q_a. The drift part can again be estimated as in the proof of Proposition 5. Put

$$\begin{aligned}
N_1 &:= X(\lfloor t_{i-1}+s \rfloor_\Delta) - X(t^{(1)}+s), &\qquad N_2 &:= X(t^{(1)}+u_1), \\
N_3 &:= X(\lfloor t_{j-1}+s \rfloor_\Delta) - X(t^{(2)}+s), &\qquad N_4 &:= X(t^{(2)}+u_2).
\end{aligned}$$

Since X is Gaussian, we obtain

$$\mathrm{Cov}_a[N_1N_2, N_3N_4] = \mathbb{E}_a[N_1N_3]\,\mathbb{E}_a[N_2N_4] + \mathbb{E}_a[N_1N_4]\,\mathbb{E}_a[N_2N_3].$$

The Lipschitz continuity of q_a at t with a Lipschitz constant of order $e^{-\delta t}$ for $\delta < -v_0(a)$ (Corollary 3) yields

$$\mathbb{E}_a[N_1N_3] \lesssim \Delta e^{-\delta|t^{(1)}-t^{(2)}|}, \qquad \mathbb{E}_a[N_2N_4] \lesssim e^{-\delta|t^{(1)}-t^{(2)}|},$$
$$\mathbb{E}_a[N_1N_4] \lesssim \Delta e^{-\delta|t^{(1)}-t^{(2)}|}, \qquad \mathbb{E}_a[N_2N_3] \lesssim \Delta e^{-\delta|t^{(1)}-t^{(2)}|}.$$

The covariance is thus of order $\Delta e^{-2\delta|t^{(1)}-t^{(2)}|}$ and the drift part can also be estimated by ΔT. □

By invoking $q_a|_{\mathbb{R}^+} \in C^{1,1}(\mathbb{R}^+)$ in the proof, one could even obtain $\mathbb{E}_a[N_1N_3] \lesssim \Delta^2 e^{-\delta|t^{(1)}-t^{(2)}|}$ and thus bound the drift part by $\Delta^2 T$.

Proposition 11. *Let X be the stationary solution of the SDDE (2.2.10) with weight function $g \in H^s([-r,0])$ and $v_0(g) < 0$. Choose s-approximating subspaces V_n and introduce the random set*

$$\mathcal{B}_g^\Delta := \{\|Q_g - \tfrac{1}{T}Q_T^\Delta\|\|(P_nQ_g|_{V_n})^{-1}\| \leq \tfrac{1}{2}\}.$$

Then the Galerkin estimator $\hat{g}_{T,n}^\Delta \in V_n$, determined by the condition

$$\langle Q_T^\Delta \hat{g}_{T,n}^\Delta, v_n\rangle = \langle b_T^\Delta, v_n\rangle, \quad \forall\, v_n \in V_n,$$

is well-defined on $\mathcal{B}_g^\Delta$ and satisfies

$$\mathbb{E}_g[\|\hat{g}_{T,n}^\Delta - g\|_{L^2}^2 \mathbf{1}_{\mathcal{B}_g^\Delta}] \lesssim n^{-2s} + n^3T^{-1} + n^4(\Delta T^{-1} + \Delta^2).$$

Proof. We apply the strategy of the proof of Proposition 10. On $\mathcal{B}_g^\Delta$ the estimator is uniquely determined by (5.1.3), and we obtain on $\mathcal{B}_g^\Delta$ the bound

$$\begin{aligned}\|\hat{g}_{T,n}^\Delta - g\| &\lesssim (2 + 2n^2n^{-2})n^{-s} + 2\|R_n\tfrac{1}{T}(Q_T^\Delta g - b_T^\Delta)\| \\ &\leq 4n^{-s} + 2\|R_nT^{-1}(Q_Tg - b_T)\| + n^2\tfrac{2}{T}(\|q_T^\Delta - q_T\|\|g\| + \|b_T^\Delta - b_T\|),\end{aligned}$$

where we used (5.3.10) in order to bound $\|R_n\| \lesssim n^2$. Putting the estimate of the variance term derived in Proposition 10 and the bounds obtained in the two preceding lemmata together, we find the announced upper bound. □

Corollary 9. *Let $s > \frac{1}{2}$, $\Delta(T) \lesssim T^{-\frac{1}{2}-\frac{1}{4s+6}}$, $n(T) \sim T^{\frac{1}{2s+3}}$ and g as in Proposition 11 with $\|g\|_s \leq S$ be given. Rescale the discretized Galerkin estimator $\hat{g}_{T,n}^\Delta$ in Proposition 11 for $\hat{g}_{T,n}^\Delta > S$ to $S\|\hat{g}_{T,n}^\Delta\|^{-1}\hat{g}_{T,n}^\Delta$ and set it to zero if $P_nQ_T^\Delta|_{V_n}$ is not invertible. Then the risk satisfies for $T \to \infty$*

$$\mathbb{E}_g[\|\hat{g}_{T,n(T)}^{\Delta(T)} - g\|_{L^2}^2]^{1/2} \lesssim T^{-\frac{s}{2s+3}},$$

which is the minimax rate for observations in continuous time.

Proof. From the preceding proposition and the risk improvement due to the rescaling we obtain

$$\mathbb{E}_g[\|\hat{g}_{T,n(T)}^{\Delta(T)} - g\|_{L^2}^2 \mathbf{1}_{\mathcal{B}_g^{\Delta(T)}}]^{1/2} \lesssim T^{-\frac{s}{2s+3}}.$$

On the complement of $\mathcal{B}_g^\Delta$ the loss is bounded by $2S$ and we only need to estimate $\mathbb{P}_g(\Omega \setminus \mathcal{B}_g^\Delta)$. The Chebyshev inequality in combination with (5.3.10), Proposition 7 and Lemma 10 yields for any $m \in \mathbb{N}$

$$\mathbb{P}_g(\|Q_g - \tfrac{1}{T}Q_T^\Delta\|\|(P_nQ_g|_{V_n})^{-1}\| > \tfrac{1}{2}) \lesssim (T^{-m} + \Delta(T)^{2m})n(T)^{4m} \lesssim T^{\frac{-2ms+m}{2s+3}}.$$

Hence, for $s > \frac{1}{2}$ the probability $\mathbb{P}_g(\Omega\setminus\mathcal{B}_g^\Delta)$ tends to zero faster than any polynomial. □

Remarks 9.

- *It is very likely that the upper bound in Proposition 11 can be improved in such a way that the term* $(\Delta T^{-1} + \Delta^2)$ *is only multiplied by* n^3*. Our proof strategy bounds* $\|R_n\|$ *by a multiple of* n^2*, whereas the variance estimation of Proposition 10 shows that the variance term is of order* $n^{3/2}$ *in* n*. If this strengthening were possible, we would obtain the more appealing result that the width* Δ *plays the same role as the level* $T^{-1/2}$ *of the noise. Heuristically, this can be understood from the fact that* q_g *and* $Q_g g$ *are Lipschitz continuous and by knowing these functions only on a grid of distance* Δ *their values can be interpolated with an error of order* Δ*. In the case of ergodic diffusions Hoffmann (1999) showed that the nonparametric estimation rate for the drift is maintained for* $\Delta T \to 0$*, which is far more restrictive than in our Gaussian SDDE case.*

- *We had achieved an estimate of order* $(T^{-1}\Delta^2)^m$ *in the moment estimation of Proposition 5. So one might wonder whether the results in the Lemmata 10 and 11 are optimal. The main error contributions in the estimates, however, are due to the bias terms and these seem to be quite tight. If we used a linear interpolation for the definition of* X^Δ*, we would obtain similar estimates. Of course, an asymptotic lower bound for observations at discrete times with* $\Delta \to 0$ *is desirable in order to find the true minimax rate for the risk. In Section 6.2 it is shown that for discrete time observations with constant uniform time step* $\Delta > 0$ *the minimax-rate is significantly slower.*

- *In the last corollary we have not given a risk bound uniformly over weight functions* g *in* $M(s, S, \delta)$*, but by retracing all the estimation steps it is not difficult to establish again the uniformity of the constants involved.*

5.5 A misspecified model

We study the Galerkin estimator $\hat{g}_{T,n}$ in the case where the true underlying weight a belongs to $M^-([-r, 0])$, but Assumption 1 is not satisfied, i.e. a might be a general finite measure, for instance $-\delta_0$, leading to an Ornstein-Uhlenbeck process. Since we cannot expect the estimator to converge in total variation norm due to the inseparability of $M([-r, 0])$ in this norm, we are interested in its convergence in the weak sense, i.e. whether $\int_{-r}^0 f(s)\hat{g}_{T,n(T)}(s)ds \to \int_{-r}^0 f(s)da(s)$ for all $f \in C([-r, 0])$ under $\mathbb{P}_a$ in some probabilistic sense.

Before proving the main result, we introduce a family of metrics $(d_a)_{a\in M^-}$ metrizing weak convergence of norm bounded sequences.

Lemma 12. *Let* $(e_i)_{i\in\mathbb{N}}$ *be an orthonormal basis of* $L^2([-r, 0])$ *and* Q_a *be a covariance operator according to Definition 4 with weight* $a \in M^-([-r, 0])$*. Then*

$$d_a(\mu, \nu) := \sum_{i=1}^{\infty} 2^{-i} \min(1, |\langle \mu - \nu, Q_a e_i \rangle|), \quad \mu, \nu \in M([-r, 0]),$$

is a metric on $M([-r, 0])$*. For norm bounded sequences* $(\mu_n) \subset M([-r, 0])$ *weak convergence* $\mu_n \xrightarrow{w} \mu \in M([-r, 0])$ *is equivalent to* $d_a(\mu_n, \mu) \to 0$*.*

Proof. That d_a is a metric is clear once the implication $d_a(\mu_1, \mu_2) = 0 \Rightarrow \mu_1 = \mu_2$ has been proved. For this it suffices to prove that $\operatorname{span}(Q_a e_i,\ i \in \mathbb{N})$ is dense in $C([-r, 0])$. Consider a measure $\mu \in M([-r, 0])$ with $\langle Q_a e_i, \mu \rangle = 0$ for all $i \in \mathbb{N}$. From the symmetry of q_a we derive $\langle Q_a \mu, e_i \rangle = 0$ for all $i \in \mathbb{N}$ and hence $Q_a \mu = 0$, because $Q_a \mu$ is an element of $C([-r, 0]) \subset L^2([-r, 0])$ and (e_i) is an orthonormal basis. The injectivity of Q_a (Proposition 3) implies $\mu = 0$, whence $\operatorname{span}(Q_a e_i,\, i \in \mathbb{N})$ lies indeed dense.

The weak convergence $\mu_n \xrightarrow{w} \mu$ implies $\int Q_a e_i d\mu_n \to \int Q_a e_i d\mu$ for all $i \in \mathbb{N}$ and the usual truncation argument shows $d_a(\mu_n, \mu) \to 0$. Conversely, $d_a(\mu_n, \mu) \to 0$ implies $\int Q_a e_i d\mu_n \to \int Q_a e_i d\mu$ for all $i \in \mathbb{N}$ and hence for all finite linear combinations. For given $f \in C([-r, 0])$ and $\varepsilon > 0$ we choose a function $h \in \operatorname{span}(Q_a e_i,\, i \in \mathbb{N})$ with $\|f - h\|_\infty < \varepsilon$. Then we obtain

$$\begin{aligned}\limsup_{n\to\infty} \Big|\int f d(\mu_n - \mu)\Big| &\le \limsup_{n\to\infty} \Big|\int h d(\mu_n - \mu)\Big| + \varepsilon(\|\mu_n\|_{TV} + \|\mu\|_{TV}) \\ &\le 2\varepsilon \sup_{n\in\mathbb{N}} \|\mu_n\|_{TV},\end{aligned}$$

so that $\mu_n \xrightarrow{w} \mu$ follows for $\sup_n \|\mu_n\|_{TV} < \infty$. □

Proposition 12. *Let a be any M^--weight and $(e_i)_{i\in\mathbb{N}}$ be an orthonormal basis of $L^2([-r, 0])$. Set $V_n := \operatorname{span}(e_1, \dots, e_n)$ and suppose that (V_n) satisfies the Bernstein inequality (5.3.7). Then the Galerkin estimator $\hat{g}_{T,n}$ from Proposition 10 satisfies for the metric d_a from Lemma 12:*

$$\mathbb{E}_a[d_a(\hat{g}_{T,n}, a)] \lesssim T^{-1/2} n^2.$$

With the choice $n(T) \sim T^{\frac{1}{2s+3}}$ from Theorem 4 the bound tends to zero for $s > \frac{1}{2}$.

Proof. Following the suitably adapted notation and the proof of Proposition 10, in particular $\|R_n\| \lesssim n^2$, and the convergence results of Chapter 4 we obtain

$$\begin{aligned}\mathbb{E}_a[d_a(\hat{g}_{T,n}, a)] &= \sum_{i=1}^{\infty} 2^{-i}\, \mathbb{E}_a[\min(1, |\langle Q_a \hat{g}_{T,n} - Q_a a, e_i\rangle|)] \\ &\le \mathbb{P}_a(\Omega \setminus \mathcal{B}_a) + \sum_{i=1}^{\infty} 2^{-i}\, \mathbb{E}_a[\mathbf{1}_{\mathcal{B}_a} |\langle (Q_a - \tfrac{1}{T} Q_T)\hat{g}_{T,n}, e_i\rangle|] \\ &\quad + \sum_{i=1}^{\infty} 2^{-i}\, \mathbb{E}_a[|\langle \tfrac{1}{T} Q_T \hat{g}_{T,n} - Q_a a, e_i\rangle|] \\ &\le \mathbb{P}_a(\Omega \setminus \mathcal{B}_a) + \mathbb{E}_a[\|q_a - \tfrac{1}{T} q_T\|_{L^2}^2]^{1/2}\, \mathbb{E}_a[\mathbf{1}_{\mathcal{B}_a} \|\hat{g}_{T,n}\|_{L^2}^2]^{1/2} \\ &\quad + \mathbb{E}_a[\|\tfrac{1}{T} b_T - Q_a a\|_{L^2}^2]^{1/2} \\ &\lesssim \mathbb{P}_a(\Omega \setminus \mathcal{B}_a) + T^{-1/2}\, \mathbb{E}_a[2\|R_n\|^2 \|\tfrac{1}{T} b_T\|_{L^2}^2]^{1/2} + T^{-1/2} \\ &\lesssim \mathbb{P}_a(\Omega \setminus \mathcal{B}_a) + T^{-1/2} n^2.\end{aligned}$$

As in the proof of Theorem 4, by Chebyshev's inequality we obtain $\mathbb{P}_a(\Omega \setminus \mathcal{B}_a) \lesssim n^2 T^{-1/2}$ so that the main statement of the proposition has been proved. If $n(T)$ is calibrated as in Theorem 4, then the bound converges to zero for $T^{\frac{-2s+1}{2s+3}} \to 0$, which is satisfied for $s > \frac{1}{2}$. □

Several questions remain open. The most important one is whether $\hat{g}_{T,n}$ remains bounded in $L^1([-r, 0])$ under $\mathbb{P}_a$, because then weak convergence would follow by Lemma 12. An L^2-bounded sequence f_n that converges weakly (in our sense) to some $\mu \in M([-r, 0])$ has a subsequence converging L^2-weakly to some $f_\infty \in L^2([-r, 0])$, so that f_∞ must be the Lebesgue density of μ. This means that in the case of weights without L^2-density a weakly convergent estimator must be L^1-bounded, but cannot remain L^2-bounded. To find such a norm bound for the Galerkin method in a non-Hilbert space is beyond the scope of the classical theory. If this L^1-boundedness could be proved, then an almost sure convergence result would show that the Galerkin estimator converges weakly to the true weight measure with probability one. This convergence would then no longer be expressed in a weight-dependent, hence unknown distance. Another goal could be to prove convergence in some $H^{-\rho}([-r,0])$-space for $\rho > \frac{1}{2}$, because $M([-r,0])$ is embedded in this space of distributions. In fact, the scalar product used in the definition of d_a induces a norm which is equivalent to the H^{-1}-norm due to the mapping properties of Q_a. The main difficulty then is the treatment of the set $\Omega \setminus \mathcal{B}_a$.

Although our result is not very strong, it shows that we can expect the Galerkin estimator to be somehow robust to misspecification, since there exists at least a metric for which it converges in the mean.

Chapter 6

Lower risk bounds

So far, we have only bounded the L^2-risk of the Galerkin estimator from above and have obtained the rate $T^{-\frac{s}{2s+3}}$, which is worse than in many other classical cases. Naturally, the question arises whether a different estimation procedure could achieve a better result. This will be answered in the negative in the first section, at least concerning the asymptotic rate, that is the above rate is really the minimax rate for our estimation problem.

Another topic of interest is whether the weight can be estimated when the trajectory is observed at discrete time points only. In Section 5.4 this problem was treated for observation points of maximal distance Δ for the asymptotic $\Delta \to 0$ and $T \to \infty$. As it turns out, the rate for $T \to \infty$ deteriorates significantly for equidistant observations of fixed distance Δ, at least to $T^{-\frac{s}{2s+6}}$. This lower bound is derived in the second section.

6.1 Continuous time observations

In order to prove the lower bound, we use the classical Assouad cube technique as presented by Härdle et al. (1998). Let $s > 0$, $S > 0$ and $\delta > 0$ be given such that $M(s, S, \delta)$ has a nonempty interior in $H^s([-r, 0])$. This condition is satisfied iff there is a weight function $g_0 \in H^s([-r, 0])$ with $\|g_0\|_s < S$ and $v_0(g) < -\delta$. For δ small enough, which is the case we are interested in, the existence of such a function is guaranteed by Lemma 1, take for instance $g_0 = -\alpha \mathbf{1}_{[-r,0]}$, $\alpha = \alpha(\delta, S, r) > 0$ small enough. On the other hand, the set $M(s, S, \delta)$ can be empty for small S, some s and large δ.

Let $(\psi_{j,k})$ be a compactly supported and s-regular wavelet basis of $L^2(\mathbb{R})$ (cf. Appendix A.3). We denote by R_j a maximal subset of $\mathbb{Z}$ with $\operatorname{supp}(\psi_{j,k}) \subset [-r, 0]$ and $\operatorname{supp}(\psi_{j,k}) \cap \operatorname{supp}(\psi_{j,k'}) = \varnothing$ for all $k, k' \in R_j$ with $k \neq k'$. The cardinality of R_j is of order $|R_j| \sim 2^j$ for $j \to \infty$ due to the compact support of the wavelets. For sign vectors $\varepsilon \in \{-1, +1\}^{R_j}$ we introduce the functions

$$g_\varepsilon := g_0 + \gamma \sum_{k \in R_j} \varepsilon_k \psi_{j,k},$$

where $\gamma = \gamma(j, T)$ is, for the moment, arbitrary, but so small that $\|g_\varepsilon\|_s \leq S$ and $v_0(g_\varepsilon) \leq -\delta$ hold. In particular, the law $\mathbb{P}_{g_\varepsilon}$ of the stationary solution with weight function g_ε is well defined. We quote the main lemma from the literature on nonparametric lower risk bounds and use it for proving our lower bound.

Lemma 13. *For* $\varepsilon = (\varepsilon_i) \in \{-1,+1\}^{R_j}$ *define* $\varepsilon^k = (\varepsilon_i^k) \in \{-1,+1\}^{R_j}$ *by*

$$\varepsilon_i^k := \begin{cases} \varepsilon_i, & \text{if } i \neq k, \\ -\varepsilon_i, & \text{if } i = k. \end{cases}$$

If there exist constants $\lambda, p > 0$ *such that for the likelihood ratio* Λ_T

$$\mathbb{P}_{g_\varepsilon}(\Lambda_T(X^{(g_{\varepsilon^k})}, X^{(g_\varepsilon)}) > e^{-\lambda}) \geq p, \quad \forall\, \varepsilon \in \{-1,+1\}^{R_j},\, k \in R_j,$$

holds, then for any $\mathcal{F}_T^X$*-measurable estimator* G_T *the following lower bound is valid:*

$$\max_{\varepsilon \in \{-1,+1\}^{R_j}} \mathbb{E}_{g_\varepsilon}[\|G_T - g_\varepsilon\|_{L^2}^2] \geq \tfrac{1}{2}|R_j|\gamma^2 e^{-\lambda} p.$$

Proof. This is Lemma 10.2 in Härdle et al. (1998) in the L^2-case, adapted to our notation. □

Theorem 5. *For* $s > 0$, $S > 0$ *and* $\delta > 0$ *such that* $M(s, S, \delta)$ *has nonempty interior in* $H^s([-r, 0])$*, the following lower bound holds for* $T \to \infty$

$$\inf_{G_T} \sup_{g \in M(s,S,\delta)} \mathbb{E}_g[\|G_T - g\|_{L^2}^2]^{1/2} \gtrsim T^{-\frac{s}{2s+3}},$$

where the infimum is taken over all $\mathcal{F}_T^X$*-measurable estimators* G_T*.*

Proof. We claim that for $2^j \sim T^{\frac{1}{2s+3}}$ (as for the upper bound) and $\gamma = c2^{-j(s+\frac{1}{2})}$ with $c > 0$ small enough the preceding lemma gives the result. Note first that $\|g_\varepsilon - g_0\|_s = \gamma 2^{j/2}\|\psi_{j,k}\|_s \sim c$ implies that for small enough c the functions g_ε remain in the Sobolev ball of radius S. Furthermore, since $\|g_\varepsilon - g_0\|_{L^2}$ converges to zero, also $v_0(g_\varepsilon) < -\delta$ will be satisfied for sufficiently large T by Theorem 1 and g_ε eventually lies in $M(s, S, \delta)$. The assertion of the theorem is then a consequence of the preceding lemma due to $|R_j|\gamma^2 \sim 2^j 2^{-j(2s+1)} \sim T^{-\frac{2s}{2s+3}}$, once universal constants λ, p have been found with

$$\mathbb{P}_{g_\varepsilon}(\log(\Lambda_T(X^{(g_{\varepsilon^k})}, X^{(g_\varepsilon)}) > -\lambda) \geq p > 0. \tag{6.1.1}$$

By Chebyshev's inequality it suffices to show that the second moment of the log-likelihood remains uniformly bounded. Since the stationary solutions are Gaussian, the laws of $X(0)$ under $\mathbb{P}_{g_\varepsilon}$ and $\mathbb{P}_{g_{\varepsilon^k}}$ are mutually absolutely continuous so that Corollary 2 in connection with Remarks 2, 2 yields for $T \geq r$

$$\begin{aligned}
&\log(\Lambda_T(X^{(g_{\varepsilon^k})}, X^{(g_\varepsilon)})(X) \\
&= \log(\Lambda_r(X^{(g_{\varepsilon^k})}, X^{(g_\varepsilon)})(X) \\
&\quad + 2\gamma\varepsilon_k \int_r^T \int_{-r}^0 X(t+s)\psi_{j,k}(s)\, ds\, dW(t) - 2\gamma^2 \int_r^T \left(\int_{-r}^0 X(t+s)\psi_{j,k}(s)\, ds\right)^2 dt \\
&=: S_1 + S_2 - S_3.
\end{aligned}$$

The term S_1 is the log-likelihood ratio between $X^{(g_{\varepsilon^k})}$ and $X^{(g_\varepsilon)}$ on $C([0, r])$. Due to the convergence of the weight functions

$$\|g_\varepsilon - g_0\|_{L^1} = \|g_{\varepsilon^k} - g_0\|_{L^1} \leq \gamma(T) 2^{|R_j(T)|/2} \to 0, \quad T \to \infty,$$

Corollary 5 yields $\mathbb{E}_{g_\varepsilon}[S_1^2] \to 0$ for $T \to \infty$.

Since S_3 is – up to a constant – the quadratic variation process in T of the martingale $(S_2, (\mathcal{F}_T)_{T \geq r})$, it suffices to establish a bound for the second moment of

S_3. We infer from the estimate $\langle Q_g\psi_{j,k},\psi_{j,k}\rangle \lesssim 2^{-2j}$ with a uniform constant for $g \in M(s,S,\delta)$ (Corollaries 3, 13) that the expected value of S_3 is bounded uniformly

$$\begin{aligned}\mathbb{E}_{g_\varepsilon}[S_3] &= 2\gamma^2 \int_r^T \int_{-r}^0 \int_{-r}^0 \mathbb{E}_{g_\varepsilon}[X(t+u)X(t+v)]\psi_{j,k}(u)\psi_{j,k}(v)\,du\,dv\,dt \\ &\leq 2\gamma^2 T\langle Q_{g_\varepsilon}\psi_{j,k},\psi_{j,k}\rangle \\ &\lesssim \gamma^2 T 2^{-2j} \sim 1.\end{aligned}$$

Proposition 7 with $\alpha = 0$ and the estimate $\|\psi_{j,k}\|_{L^1} \sim 2^{-j/2}$ yield for the variance term

$$\begin{aligned}\mathrm{Var}_{g_\varepsilon}[S_3] &\leq 4\gamma^4\,\mathbb{E}_{g_\varepsilon}[\langle (Q_T - TQ_{g_\varepsilon})\psi_{j,k},\psi_{j,k}\rangle^2] \\ &\leq 4\gamma^4\,\mathbb{E}_{g_\varepsilon}[\|q_T - Tq_{g_\varepsilon}\|_\infty^2]\|\psi_{j,k}\|_{L^1}^4 \\ &\lesssim \gamma^4 T 2^{-2j} \sim \gamma^2 \to 0.\end{aligned}$$

Again, the constant involved can be chosen uniformly for $g \in M(s,S,\delta)$ by Proposition 7 and the Chebyshev inequality yields (6.1.1) with constants $\lambda > 0$, $p > 0$ independent of ε, k and γ:

$$\begin{aligned}\mathbb{P}_{g_\varepsilon}(\log(\Lambda_T(X^{(g_{\varepsilon k})}, X^{(g_\varepsilon)}) > -\lambda) &\geq 1 - \mathbb{P}_{g_\varepsilon}(|\log(\Lambda_T(X^{(g_{\varepsilon k})}, X^{(g_\varepsilon)})| > \lambda) \\ &\geq 1 - C_1(\mathbb{E}[S_1^2] + \mathbb{E}[(S_2 - S_3)^2]) \\ &\geq 1 - C_2,\end{aligned}$$

with constants C_1, $C_2 > 0$, where C_2 can be made arbitrarily small by choosing $c = \gamma 2^{j(s+\frac{1}{2})}$ sufficiently small. □

Corollary 10. *For $s > \frac{1}{2}$, $S > 0$ and $\delta > 0$ such that $M(s,S,\delta)$ has nonempty interior in $H^s([-r,0])$, the minimax rate of convergence for estimating the unknown weight function from the continuous observation of a stationary solution up to time T is given by*

$$\inf_{G_T} \sup_{g \in M(s,S,\delta)} \mathbb{E}_g[\|G_T - g\|_{L^2}^2]^{1/2} \sim T^{-\frac{s}{2s+3}}, \quad T \to \infty,$$

where the infimum is taken over all $\mathcal{F}_T^X$-measurable estimators G_T.

Proof. This is the joint result of Theorem 4 and Theorem 5. Only observe that the Galerkin estimator $\hat{g}_{T-r,n}$ is $\sigma(X(t), -r \leq t \leq T-r)$-measurable, but by stationarity equals in law the corresponding estimator based on observations of $(X(t), 0 \leq t \leq T)$; use $T - r \sim T$ to conclude. □

Remarks 10.

- *In Theorem 7 we shall see that the condition $s > \frac{1}{2}$ can be relaxed when logarithmic terms are neglected, because the adaptive estimator attains its rate for all $s > 0$. By constructing a non-adaptive linear estimator of $Q_a a$ in $H^2([-r,0])$ and proceeding as in Chapter 7 it is even possible to get rid of the logarithmic factor.*

- *Note that we have shown the lower bound even for weight functions in the Besov space $B^s_{\infty,\infty}([-r,0])$ due to $\|g_\varepsilon - g_0\|_{s,\infty,\infty} \lesssim \gamma 2^{j(s+\frac{1}{2})} \sim 1$.*

6.2 Discrete time observations (Δ fixed)

Suppose now that we observe the stationary solution X of the SDDE (2.2.10) at discrete time points $0, \Delta, 2\Delta, \dots, N\Delta$. For a fixed width Δ we find an asymptotic lower risk bound for $N \to \infty$. Due to $N = \Delta^{-1}T$ the obtained rate $N^{-\frac{s}{2s+6}}$ is worse than the minimax rate $T^{-\frac{s}{2s+3}}$ for continuous time observations.

It should be emphasized that only a lower bound is known and that a minimax rate might be slower or that the weight is not even identifiable from discrete time observations. In fact, the observer only sees a discrete time Gaussian process evolving with (auto-)covariances $(q_a(i\Delta))_{i\in\mathbb{N}_0}$ and the weight a will be identifiable if $a \mapsto (q_a(i\Delta))_{i\in\mathbb{N}_0}$ is an injective map on the considered set of parameters. This is why we shall state the result locally for a small nonparametric neighbourhood of a weight function, which could possibly be chosen in such a way that the identifiability property is satisfied. We start with a lemma on Gaussian likelihood functions preparing the grounds for the main theorem. For the theory of Hilbert-Schmidt operators used throughout this section we refer to Dunford and Schwartz (1963, Section XI.6). Let us only mention that the Hilbert-Schmidt norm of a symmetric operator S on $\mathbb{R}^d$ is given in terms of its eigenvalues $(\lambda_i)_{1\le i\le d}$ (counted according to their multiplicity) by $(\sum_{i=1}^d \lambda_i^2)^{1/2}$.

Lemma 14. *For two centred d-dimensional Gaussian laws $\mathbb{P}_1$ and $\mathbb{P}_2$ with regular covariance matrices $\mathcal{Q}_1$ and $\mathcal{Q}_2$, respectively, introduce the symmetric matrix $L := \mathcal{Q}_2^{1/2}\mathcal{Q}_1^{-1}\mathcal{Q}_2^{1/2} - \mathrm{Id}$ and assume that all of its eigenvalues λ_i satisfy $\lambda_i \ge -\frac{1}{2}$. Then the following estimate in terms of the Hilbert-Schmidt norm $\|L\|_{HS}$ is satisfied by the likelihood function $\Lambda_{1,2} := \frac{d\,\mathbb{P}_1}{d\,\mathbb{P}_2}$:*

$$\mathbb{P}_2(\Lambda_{1,2} \ge \exp(-\tfrac{1}{2}\|L\|_{HS}^2 - \|L\|_{HS})) \ge \tfrac{1}{4}.$$

Proof. First, we express the law of the log-likelihood function under $\mathbb{P}_2$ in terms of L using the product rule for determinants:

$$\begin{aligned}
&\log(\Lambda_{1,2}(x))\\
&= \tfrac{1}{2}\log(\det(\mathcal{Q}_1^{-1}\mathcal{Q}_2)) - \tfrac{1}{2}\langle(\mathcal{Q}_1^{-1} - \mathcal{Q}_2^{-1})x, x\rangle \quad \text{under } \mathbb{P}_2\\
&= \tfrac{1}{2}\log(\det(\mathcal{Q}_2^{1/2}\mathcal{Q}_1^{-1}\mathcal{Q}_2^{1/2})) - \tfrac{1}{2}\langle(\mathcal{Q}_1^{-1} - \mathcal{Q}_2^{-1})\mathcal{Q}_2^{1/2}x, \mathcal{Q}_2^{1/2}x\rangle \quad \text{under } N(0,\mathrm{Id})\\
&= \tfrac{1}{2}\log(\det(\mathrm{Id} + L)) - \tfrac{1}{2}\langle Lx, x\rangle \quad \text{under } N(0,\mathrm{Id}).
\end{aligned}$$

Being a symmetric matrix, L has an orthonormal basis $(e_1, \dots, e_d)$ of eigenvectors with corresponding eigenvalues $\lambda_1, \dots, \lambda_d$. From the inequality $\log(1+\lambda) \ge \lambda - \lambda^2$ for $\lambda \ge -\frac{1}{2}$ and the assumption on the eigenvalues (λ_i) follows for all $x = \sum_{i=1}^d x_i e_i \in \mathbb{R}^d$

$$\log(\det(\mathrm{Id} + L)) - \langle Lx, x\rangle = \sum_{i=1}^d (\log(1+\lambda_i) - \lambda_i x_i^2) \ge \sum_{i=1}^d (\lambda_i - \lambda_i^2 - \lambda_i x_i^2) =: l(x).$$

For an $N(0,\mathrm{Id})$-distributed random vector Z in $\mathbb{R}^d$ we find

$$\mathbb{E}[\tfrac{1}{2}l(Z)] = -\tfrac{1}{2}\|L\|_{HS}^2 \text{ and } \mathrm{Var}[\tfrac{1}{2}l(Z)] = \tfrac{3}{4}\|L\|_{HS}^2.$$

By Chebyshev's inequality we conclude

$$\begin{aligned}
\mathbb{P}_2(\Lambda_{1,2} \ge \exp(-\tfrac{1}{2}\|L\|_{HS}^2 - \|L\|_{HS})) &\ge 1 - \mathbb{P}(\tfrac{1}{2}l(Z) + \tfrac{1}{2}\|L\|_{HS}^2 < -\|L\|_{HS})\\
&\ge 1 - \frac{\frac{3}{4}\|L\|_{HS}^2}{\|L\|_{HS}^2} = \frac{1}{4}.
\end{aligned}$$

□

Theorem 6. *Let $s > 0$, $\delta > 0$ and $g_0 \in H^s([-r,0])$ with $v_0(g_0) < -\delta$ be given. Then the following local risk lower bound for equidistant discrete time observations holds for all fixed $\Delta > 0$ and $\varepsilon > 0$:*

$$\inf_{G_N^\Delta} \sup_{\substack{g\in H^s([-r,0])\\ v_0(g)\le -\delta,\, \|g-g_0\|_s\le\varepsilon}} \mathbb{E}_g[\|g-G_N^\Delta\|_{L^2}^2]^{1/2} \gtrsim N^{-\frac{s}{2s+6}},$$

where the infimum is taken over all $\sigma(X(0),X(\Delta),\dots,X(N\Delta))$-measurable estimators G_N^Δ of g.

Proof. The proof follows the same ideas as the one for continuous time observations in the last section. We choose $S := \|g_0\|_s + \varepsilon$, use an $(s\vee 3)$-regular wavelet basis $(\psi_{j,k})$, define R_j as before and put

$$g_\varepsilon := g_0 + \gamma \sum_{k\in R_j} \varepsilon_k \psi_{j,k}.$$

This time we ensure by the right choice of $\gamma = \gamma(j,N)$ that $\|g_\varepsilon - g_0\|_s \le \varepsilon$ and $v_0(g_\varepsilon) \le -\delta$ is satisfied; hence $g \in M(s,S,\delta)$ holds. This means that we take $\gamma = c2^{-j(s+\frac{1}{2})}$ with a sufficiently small constant $c > 0$. Furthermore, we put $2^j \sim N^{\frac{1}{2s+6}}$. Then, in terms of the likelihood ratio

$$\Lambda_N^\Delta(X^{(g_1)},X^{(g_2)}) = \mathbb{E}_{g_2}\left[\frac{d\,\mathbb{P}_{g_1}}{d\,\mathbb{P}_{g_2}} \,\middle|\, \sigma(X(0),\dots,X(N\Delta))\right],$$

Assouad's Lemma 13 with adapted notation yields the statement of the theorem as soon as universal constants λ, p have been found with

$$\mathbb{P}_{g_\varepsilon}(\Lambda_N^\Delta(X^{(g_{\varepsilon^k})},X^{(g_\varepsilon)}) > e^{-\lambda}) \ge p > 0; \tag{6.2.2}$$

for our choice of γ and j yields

$$\tfrac{1}{2}|R_j|\gamma^2 e^{-\lambda} p \sim 2^j 2^{-j(2s+1)} \sim N^{-\frac{2s}{2s+6}}.$$

The stationary solution X is a nondegenerate Gaussian process whence Λ_N^Δ in (6.2.2) is the likelihood between finite-dimensional Gaussian vectors and we can apply Lemma 14. In the notation of this lemma we consider the covariance matrices $\mathcal{Q}_1 = (q_1((i-j)\Delta))_{1\le i,j\le N}$ with $q_1(t) := q_{g_{\varepsilon^k}}(t)$ and $\mathcal{Q}_2 = (q_2((i-j)\Delta))_{1\le i,j\le N}$ with $q_2(t) := q_{g_\varepsilon}(t)$ and the problem is reduced to bounding the Hilbert-Schmidt norm of the operator L, provided all eigenvalues of L are larger than $-\frac{1}{2}$. By the ideal property of Hilbert-Schmidt operators Dunford and Schwartz (1963, Cor. XI.6.5) we obtain the estimate

$$\|L\|_{HS} = \|\mathcal{Q}_2^{1/2}\mathcal{Q}_1^{-1}(\mathcal{Q}_2-\mathcal{Q}_1)\mathcal{Q}_2^{-1/2}\|_{HS} \le \|\mathcal{Q}_2^{1/2}\|\|\mathcal{Q}_1^{-1}\|\|\mathcal{Q}_2-\mathcal{Q}_1\|_{HS}\|\mathcal{Q}_2^{-1/2}\|. \tag{6.2.3}$$

We prove next that the operator norms $\|\mathcal{Q}_2^{1/2}\|$, $\|\mathcal{Q}_1^{-1}\|$ and $\|\mathcal{Q}_2^{-1/2}\|$ are uniformly bounded by showing that the eigenvalues of the symmetric and positive definite covariance matrix $\mathcal{Q}_g = (q_g((i-j)\Delta))_{i,j}$ are bounded away from zero and from infinity, uniformly for weight functions g in $M(s,S,\delta)$. In the last step we will establish the fact that the Hilbert-Schmidt norm $\|\mathcal{Q}_1-\mathcal{Q}_2\|_{HS}$ is uniformly of order $\gamma N^{1/2}2^{-5j/2}$, hence also uniformly bounded, which then proves the assertion.

Denote by $\mathcal{Q}$ any covariance matrix $\mathcal{Q}_g$ with $g \in M(s,S,\delta)$ arbitrary. For bounding the eigenvalues of $\mathcal{Q}$ from above and below, we use the representation of $\mathcal{Q}$ by virtue of the spectral density $\hat{q} = |\chi|^{-2}(i\bullet)$. For $x = (x_0,\dots,x_N) \in \mathbb{R}^{N+1}$ we set

$f_x(t) = \sum_{k=0}^N x_k e^{ikt}$ and calculate (the single arguments will be given afterwards)

$$\begin{aligned}
\langle \mathcal{Q}x, x\rangle_{\mathbb{R}^{N+1}} &= \sum_{k,l=0}^{N} q((k-l)\Delta)x_k x_l \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\sum_{k=-2N}^{2N} q(k\Delta)e^{-ikt}\right)|f_x(t)|^2\,dt \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\sum_{k=-2N}^{2N} \mathcal{F}^{-1}(|\chi(i\bullet)|^{-2})(k\Delta)e^{-ikt}\right)|f_x(t)|^2\,dt \\
&= \frac{1}{4\pi^2}\int_{-\infty}^{\infty}|\chi(i\xi)|^{-2}\sum_{k=-2N}^{2N}\int_{-\pi}^{\pi} e^{-ikt}|f_x(t)|^2\,dt\,e^{ik\Delta\xi}\,d\xi \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{|f_x(\Delta\xi)|^2}{|\chi(i\xi)|^2}\,d\xi
\end{aligned}$$

In the second line the orthogonality of the exponentials was used. The integrals in the fourth line may be interchanged due to the uniform boundedness in t and the integrability in ξ of the integrand. The last line follows from the observation that the Fourier decomposition of the trigonometric function $|f_x|^2$ returns $|f_x|^2$.

Since the Euclidean norm of x satisfies $|x|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}|f_x(t)|^2 dt$, we are now in the position to bound the smallest eigenvalue of $\mathcal{Q}$ by

$$\begin{aligned}
\inf_{|x|=1}\langle \mathcal{Q}x, x\rangle &\geq \tfrac{1}{2\pi}\inf_{|x|=1}\int_{-\pi}^{\pi}\frac{|f_x(\xi)|^2}{|\chi(i\Delta^{-1}\xi)|^2}\Delta^{-1}\,d\xi \\
&\geq \Delta^{-1}\min_{|\xi|\leq\pi}(|\chi(i\Delta^{-1}\xi)|^{-2}) > 0.
\end{aligned}$$

Hence, by Lemma 2 and Corollary 1 there is a positive constant which bounds the smallest eigenvalue of $\mathcal{Q}$ from below for all weights in $M(s,S,\delta) \subset M(S,\delta)$, independent of N. In order to establish a uniform upper bound for the eigenvalues of $\mathcal{Q}$, recall the uniform estimate $|\chi_g(i\xi)| \gtrsim (1+\xi^2)^{1/2}$, $\xi \in \mathbb{R}$, for $g \in M(s,S,\delta)$ from the proof of Proposition 1, which yields

$$\begin{aligned}
\sup_{|x|=1}\langle \mathcal{Q}x, x\rangle &= \tfrac{1}{2\pi}\sup_{|x|=1}\sum_{k\in\mathbb{Z}}\int_{2k\pi}^{2(k+1)\pi}\frac{|f_x(\xi)|^2}{|\chi(i\Delta^{-1}\xi)|^2}\Delta^{-1}\,d\xi \\
&\leq \Delta^{-1}\sum_{k\in\mathbb{Z}}\max_{\xi\in[2k\pi,2(k+1)\pi]}|\chi(i\Delta^{-1}\xi)|^{-2} \\
&\lesssim \sum_{k\in\mathbb{Z}}(1+k^2)^{-1} < \infty.
\end{aligned}$$

Hence, the eigenvalues of $\mathcal{Q}$ are indeed bounded away from zero and bounded from above uniformly on $M(s,S,\delta)$, so that (6.2.3) simplifies to the uniform estimate

$$\|L\|_{HS} \lesssim \|\mathcal{Q}_2 - \mathcal{Q}_1\|_{HS}.$$

The Hilbert-Schmidt norm of the matrix $\mathcal{Q}_2 - \mathcal{Q}_1$ can be calculated componentwise Dunford and Schwartz (1963, Cor. XI.6.3) such that

$$\|\mathcal{Q}_2 - \mathcal{Q}_1\|_{HS}^2 = \sum_{i,j=0}^{N}(q_2-q_1)^2((i-j)\Delta) \leq 2(N+1)\sum_{i=0}^{N}(q_2-q_1)^2(i\Delta)$$

holds. We use the bound (2.3.19) in the proof of Proposition 1 with δ replaced by $\frac{\delta}{2}$ to establish an exponential decay

$$
\begin{aligned}
&\sup_{t\geq 0} e^{\frac{\delta}{2}t}|q_2(t)-q_1(t)| \\
&\leq \frac{1}{2\pi}\int_{-\infty}^{\infty}|\hat{q}_2(\xi+i\tfrac{\delta}{2})-\hat{q}_1(\xi+i\tfrac{\delta}{2})|\,d\xi \\
&\lesssim \int_{-\infty}^{\infty} M^3(\xi)\big(|\chi_1(-i\xi+\tfrac{\delta}{2})-\chi_2(-i\xi+\tfrac{\delta}{2})|+|\chi_1(i\xi-\tfrac{\delta}{2})-\chi_2(i\xi-\tfrac{\delta}{2})|\big)\,d\xi \\
&\lesssim \int_{-\infty}^{\infty}(1+\xi^2)^{-3/2}\gamma(|\widehat{\psi_{jk}}(-\xi-i\tfrac{\delta}{2})|+|\widehat{\psi_{jk}}(\xi+i\tfrac{\delta}{2})|)\,d\xi \\
&= \gamma\int_{-\infty}^{\infty}(1+\xi^2)^{-3/2}2^{-j/2}(|\hat{\psi}(-2^{-j}\xi-i2^{-j-1}\delta)|+|\hat{\psi}(2^{-j}\xi+i2^{-j-1}\delta)|)\,d\xi \\
&= \gamma 2^{-5j/2}\int_{-\infty}^{\infty}(2^{-2j}+\xi^2)^{-3/2}(|\hat{\psi}(-\xi-i2^{-j-1}\delta)|+|\hat{\psi}(\xi+i2^{-j-1}\delta)|)\,d\xi.
\end{aligned}
$$

We claim that the last integral is bounded uniformly for $j\to\infty$. Observe first by the Cauchy-Schwarz inequality and the Plancherel theorem

$$
\begin{aligned}
&\int_{|\xi|>1}(2^{-2j}+\xi^2)^{-3/2}(|\hat{\psi}(-\xi-i2^{-j-1}\delta)|+|\hat{\psi}(\xi+i2^{-j-1}\delta)|)\,d\xi \\
&\leq \left(\int_{|\xi|>1}|\xi|^{-6}\,d\xi\right)^{1/2}(\|\hat{\psi}(-\bullet-i2^{-j-1}\delta)\|_{L^2}+\|\hat{\psi}(-\bullet-i2^{-j-1}\delta)\|_{L^2}) \\
&\lesssim \|E_{2^{-j-1}\delta}(\psi)\|_{L^2}+\|E_{-2^{-j-1}\delta}(\psi)\|_{L^2},
\end{aligned}
$$

which is uniformly bounded, because ψ has compact support. Secondly, the compact support and the vanishing moments of ψ up to the third order allow to define a compactly supported third antiderivative

$$
\psi^{(-3)}(x)=\int_{-\infty}^{x}\int_{-\infty}^{\xi}\int_{-\infty}^{\eta}\psi(\zeta)\,d\zeta\,d\eta\,d\xi,\quad x\in\mathbb{R}.
$$

By partial integration of the Fourier transform we obtain on $[-1,1]$

$$
\begin{aligned}
&\int_{-1}^{1}(2^{-2j}+\xi^2)^{-3/2}(|\hat{\psi}(-\xi-i2^{-j-1}\delta)|+|\hat{\psi}(\xi+i2^{-j-1}\delta)|)\,d\xi \\
&\leq \int_{-1}^{1}|\xi+i2^{-j}|^{-3}(|\widehat{\psi^{(-3)}}(-\xi-i2^{-j-1}\delta)|+|\widehat{\psi^{(-3)}}(\xi+i2^{-j-1}\delta)|)|\xi+i2^{-j-1}\delta|^3\,d\xi \\
&\lesssim \int_{-1}^{1}(|\widehat{\psi^{(-3)}}(-\xi-i2^{-j-1}\delta)|+|\widehat{\psi^{(-3)}}(\xi+i2^{-j-1}\delta)|)\,d\xi \\
&\leq \|\widehat{\psi^{(-3)}}(-\bullet-i2^{-j-1}\delta)\|_{L^2}+\|\widehat{\psi^{(-3)}}(-\bullet-i2^{-j-1}\delta)\|_{L^2} \\
&= \|\mathbb{E}_{2^{-j-1}\delta}\,\psi^{(-3)}\|_{L^2}+\|\mathbb{E}_{-2^{-j-1}\delta}\,\psi^{(-3)}\|_{L^2},
\end{aligned}
$$

which is, for the same reasons as above for ψ, uniformly bounded.

Joining the estimates, we can finally bound the Hilbert-Schmidt norm of L:

$$
\|L\|_{HS}^2\lesssim N\sum_{k=0}^{N}(q_1-q_2)(k\Delta)^2\lesssim N\gamma^2 2^{-5j}\sum_{k=0}^{\infty}e^{-k\delta\Delta/2}\sim N\gamma^2 2^{-5j}\sim c^2.
$$

The estimate holds uniformly for $g\in M(s,S,\delta)$. We obtain the bound $\|L\|_{HS}\leq\frac{1}{2}$ by the right choice of c and hence $|\lambda_i|\leq\frac{1}{2}$ for all eigenvalues λ_i of L. Lemma 14 then establishes (6.2.2) uniformly and an application of Lemma 13 concludes the proof. $\square$

Remarks 11.

- *The technique employed in the proof would yield the same lower bound for N independent observations of the value $X(0)$. Loosely speaking, we used the fact that the mapping $g \mapsto q_g$ is continuous with respect to the $B^{-3}_{\infty,\infty}$-norm and the sup-norm in connection with $\|\psi_{jk}\|_{-3,\infty,\infty} = 2^{-5j/2}$. This leads to a value of at least $\frac{5}{2}$ for the degree of ill-posedness involved and explains the appearance of the value 6 in the rate as the sum of twice the degree of ill-posedness plus the effect of approximate white noise. A way to improve, i.e. to increase the lower risk bound could be to find (maybe even nonlocal) perturbations g_ε which leave the values of the covariance function at the points $i\Delta$ for $i = 0, \dots, I$ invariant with $I \to \infty$ as $N, j \to \infty$. This would yield an exponential decay of $\|L\|_{HS}$ in terms of I.*

- *Concerning the question of identifiability, we are faced with the problem of identifying the weight a from the sequence $(q_a(k\Delta))_{k\in\mathbb{N}_0}$, which is known after an infinitely long observation time. Since q_a solves the deterministic delay equation (2.1.1), this can be considered as an analytical problem in the theory of deterministic delay differential equations, where already problems appear in the case of solutions known on continuous time intervals Verduyn Lunel (2000). If the observations are not equidistant, then the above lower bound and also the identifiability problem might change completely. In the case of ergodic autonomous diffusions the drift and the diffusion coefficient can be identified from discrete time observations, because the transition operator of the associated Markov semigroup is identifiable for $T \to \infty$.*

Intuitively, the possibilities to increase the lower bound seem to be more promising, which would mean that the whole estimation problem is even more ill-posed. As a general guideline, the estimation should hence only be performed for high frequency data. It should be stressed that even in the case of parametric estimation for affine SDDEs yet no result exists for discrete time observations.

Chapter 7

Adaptive estimation

The Galerkin estimator has two main drawbacks. Firstly, it requires an a priori knowledge or guess of the regularity s of the true weight function for the right choice of the approximation space V_n. Secondly, it is not well suited for functions that have a strongly varying pointwise regularity. For this kind of functions the approximation spaces should be chosen adaptively, that is nonlinear methods are recommendable. In our case the locations of the more irregular regions of the weight function are not known, but wavelet thresholding or adaptive kernel estimation methods successfully try to locate them approximately from the known data.

An approach based on wavelet thresholding is presented in the first section. It is more involved than the Galerkin estimator, but attains the optimal rate of convergence up to a logarithmic factor even for the wider class of Besov spaces, which are a means to characterize spatially inhomogeneous regularity. The rate-optimality of the adaptive estimator with respect to an L^2-risk function is derived from the lower bound of the second section.

7.1 The thresholding estimator

The main idea for the construction of an adaptive estimator of the weight is to work in the image space of the covariance operator Q, which is the Sobolev space $H^2([-r,0])$ for the generalized L^2-domain $\mathcal{W}^0_{2,2}$. The first step consists of estimating adaptively the function $Q_a a$ from the data b_T such that the H^2-norm of the error is small. In the second step we find a good approximation $\widehat{Q}_T$ of the covariance operator Q_a and apply $\widehat{Q}_T^{-1}$ to the estimate of $Q_a a$.

The adaptive estimate $\hat{y}_T$ of $Q_a a$ is obtained by the so-called hard thresholding algorithm. The function $\frac{1}{T}b_T$ can be regarded as a noisy observation of $Q_a a$ (Corollary 8). We calculate the wavelet coefficients (Y_λ) of $\frac{1}{T}b_T$ and we only use those coefficients that are larger than a certain threshold level to construct a smoothed version $\hat{y}_T$ of $\frac{1}{T}b_T$. The reason for this procedure to work is that small coefficients of the observation $\frac{1}{T}b_T$ are with high probability due to small or vanishing coefficients of the true function $Q_a a$. Neglecting them does not very much increase the bias, but keeps the variance small. At points of less regularity of $Q_a a$ the wavelet coefficients will be large even for higher frequencies 2^j, hence more wavelets are used for the estimator than in regions of slow variation.

Definition 10. *Let $s_{max} > 2$ be fixed. With b_T from Definition 6 introduce for any* *multi-index λ the coefficient*

$$Y_{\lambda,T} := \langle \tfrac{1}{T} b_T, \psi_\lambda \rangle,$$

where $(\psi_\lambda)_\lambda$ is a compactly supported s_{max}-regular wavelet basis in $L^2([-r,0])$. Define the hard thresholding estimator

$$\hat{y}_T := \hat{y}_{T,J(T),\kappa(T)} := \sum_{|\lambda|\le J(T)} \left(Y_{\lambda,T} \mathbf{1}_{|Y_{\lambda,T}|>\kappa_\lambda(T)} \right) \psi_\lambda$$

for a certain resolution level $J(T)$ and thresholds $\kappa(T) := (\kappa_\lambda(T))$.

The next definition introduces the classes of weights for which the minimax rates will be considered (cf. Definition 5). Furthermore, two constants will be needed for the determination of the threshold level.

Definition 11. *For $s > 0$, $S > 0$, $p > 1$ and $\delta > 0$ set*

$$M(s,p,S,\delta) := \{a \in \mathcal{W}^s_{p,\infty} \mid \|a\|_{s,p,\infty} \le S,\ v_0(a) \le -\delta\}.$$

Denote by $\alpha(S,\delta)$ the constant $K^{-1}S^{-1}$ with K from Proposition 6, chosen uniformly for weights in $M(S,\delta)$. Denote by $\beta(S,\delta)$ the constant K^{-1} with K from Lemma 8, also chosen uniformly for weights in $M(S,\delta)$.

We now prove the main proposition on adaptive estimation of $Q_a a$. In the case of L^2-loss, this is fairly classical. An abstract approach covering much more general situations has been developed by Kerkyacharian and Picard (2000), and we largely follow their ideas. A specification of their results to our situation would require an amount of work similar to the following self-contained presentation.

Proposition 13. *Let $s \in (0, s_{max} - 2]$, $S > 0$, $\max(\frac{6}{2s+3}, 1) < p < \infty$ and $\delta > 0$ be given. Set $2^{J(T)} \sim T^{1/2}$ and $\kappa_\lambda(T) = \kappa 2^{-j} T^{-1/2} \log(T)$ with some $\kappa \ge \frac{48}{\alpha(S,\delta)}$ and $\kappa \ge \frac{24}{\beta(S,\delta)}$. Then we obtain the following asymptotic estimate for the estimator $\hat{y}_T$ from Definition 10:*

$$\sup_{a\in M(s,p,S,\delta)} \mathbb{E}_a[\|\hat{y}_T - Q_a a\|^2_{H^2}]^{1/2} \lesssim \left(\frac{T}{(\log T)^2} \right)^{-\frac{s}{2s+3}}.$$

Remark 8. *Although the convergence rate depends on s, the estimator $\hat{y}_T$ is completely independent of s. It depends however on S and δ via the thresholding level κ and on s_{max} by the choice of the wavelet basis.*

Proof. For the sake of brevity we shall suppress the T-dependence of the quantities, i.e. rather write Y_λ, J or κ_λ. We introduce the true coefficients (Z_λ) and error coefficients (E_λ)

$$Z_\lambda := \langle Q_a a, \psi_\lambda \rangle, \quad E_\lambda := Y_\lambda - Z_\lambda = \langle \tfrac{1}{T} b_T - Q_a a, \psi_\lambda \rangle.$$

For E_λ we obtain from Proposition 6 and Lemma 8 by the Cauchy-Schwarz inequality the uniform exponential moment estimate

$$\begin{aligned}
\sup_\lambda \mathbb{E}_a[\exp(\tfrac{6}{\kappa} T^{1/2} 2^{|\lambda|} |E_\lambda|)] &\le \sup_\lambda \mathbb{E}_a[\exp(\tfrac{12}{\kappa} T^{1/2} 2^{|\lambda|} |\langle (\tfrac{1}{T} Q_T - Q_a) a, \psi_\lambda \rangle|)]^{1/2} \bullet \\
&\quad \bullet \sup_\lambda \mathbb{E}_a[\exp(\tfrac{12}{\kappa} T^{-1/2} 2^{|\lambda|} |\langle b_T - Q_T a, \psi_\lambda \rangle|)]^{1/2} \\
&\le \sup_\lambda \mathbb{E}_a[\cosh(\tfrac{1}{2} \alpha(S,\delta) T^{1/2} 2^{|\lambda|} \langle (\tfrac{1}{T} Q_T - Q_a) a, \psi_\lambda \rangle)]^{1/2} \\
&\quad \bullet \sup_\lambda \mathbb{E}_a[\exp(\tfrac{1}{2} \beta(S,\delta) T^{-1/2} 2^{|\lambda|} |\langle b_T - Q_T a, \psi_\lambda \rangle|)]^{1/2} \\
&\le C < \infty. \qquad (7.1.1)
\end{aligned}$$

Proposition 6 and Lemma 8 yield a uniform constant C for all weights $a \in M(s,p,S,\delta) \subset M(S,\delta)$.

Starting with the error estimate, we obtain from the characterisation of the space $H^2([-r,0])$ by 2-regular wavelets (Appendix A.3)

$$\mathbb{E}_a[\|\hat{y}_T - Q_a\|_{H^2}^2] \sim \mathbb{E}_a\left[\sum_{|\lambda|\le J} 2^{4|\lambda|}\left(Y_\lambda \mathbf{1}_{|Y_\lambda|>\kappa_\lambda} - Z_\lambda\right)^2\right] + \sum_{|\lambda|>J} 2^{4|\lambda|} Z_\lambda^2.$$

The second term can be dealt with by linear approximation theory. By Jackson's inequality in $H^2([-r,0])$ (Corollary 11), the Besov space embeddings (A.2.3) and the restriction $p > \frac{6}{2s+3}$ we find with a uniform constant for $a \in M(s,p,S,\delta)$

$$\begin{aligned}
\sum_{|\lambda|>J} 2^{4|\lambda|} Z_\lambda^2 &\sim \|(\mathrm{Id} - P_J) Q_a a\|_{H^2}^2 \\
&\lesssim 2^{-2J\frac{2s}{2s+3}} \|Q_a a\|^2_{H^{\frac{2s}{2s+3}+2}} \\
&\lesssim T^{-\frac{2s}{2s+3}} \|Q_a a\|^2_{B_{p,\infty}^{\frac{2s}{2s+3}+\max(\frac12-\frac1p,0)+2}} \\
&\le T^{-\frac{2s}{2s+3}} \|Q_a a\|^2_{B_{p,\infty}^{\frac{2s}{2s+3}+\frac{s}{3}+2}} \\
&\lesssim T^{-\frac{2s}{2s+3}} \|Q_a a\|^2_{B_{p,\infty}^{s+2}}.
\end{aligned}$$

Due to $Q_a : \mathcal{W}^s_{p,\infty} \to B^{s+2}_{p,\infty}$ isomorphically (Theorem 2) with uniform constants (Corollary 7), this second term is of order $T^{-\frac{2s}{2s+3}}$ uniformly over $M(s,p,S,\delta)$.

The summands of the first term can be split according to the four cases whether thresholding takes place or not and whether the true coefficient is large or not:

$$\begin{aligned}
\left(Y_\lambda \mathbf{1}_{|Y_\lambda|>\kappa_\lambda} - Z_\lambda\right)^2 &= (Y_\lambda - Z_\lambda)^2 \mathbf{1}_{|Y_\lambda|>\kappa_\lambda} + Z_\lambda^2 \mathbf{1}_{|Y_\lambda|\le\kappa_\lambda} \\
&= E_\lambda^2 \mathbf{1}_{\substack{|Y_\lambda|>\kappa_\lambda \\ |Z_\lambda|\le\kappa_\lambda/2}} + E_\lambda^2 \mathbf{1}_{\substack{|Y_\lambda|>\kappa_\lambda \\ |Z_\lambda|>\kappa_\lambda/2}} + Z_\lambda^2 \mathbf{1}_{\substack{|Y_\lambda|\le\kappa_\lambda \\ |Z_\lambda|>2\kappa_\lambda}} + Z_\lambda^2 \mathbf{1}_{\substack{|Y_\lambda|\le\kappa_\lambda \\ |Z_\lambda|\le2\kappa_\lambda}} \\
&\le E_\lambda^2 \mathbf{1}_{|E_\lambda|>\kappa_\lambda/2} + E_\lambda^2 \mathbf{1}_{|Z_\lambda|>\kappa_\lambda/2} + Z_\lambda^2 \mathbf{1}_{|E_\lambda|>\kappa_\lambda} + Z_\lambda^2 \mathbf{1}_{|Z_\lambda|\le 2\kappa_\lambda} \\
&=: S_1(\lambda) + S_2(\lambda) + S_3(\lambda) + S_4(\lambda).
\end{aligned}$$

By the Cauchy-Schwarz inequality, the exponential moment property of E_λ (7.1.1) and $x^4 \lesssim e^x$, $x \ge 0$, we obtain a fast decay for the sum over $\mathbb{E}_a[S_1(\lambda)]$:

$$\begin{aligned}
\sum_{|\lambda|\le J} 2^{4|\lambda|}\,\mathbb{E}_a[S_1(\lambda)] &\le \sum_{|\lambda|\le J} 2^{4|\lambda|}\,\mathbb{P}_a(|E_\lambda| > \tfrac{\kappa_\lambda}{2})^{1/2}\,\mathbb{E}_a[E_\lambda^4]^{1/2} \\
&\le \sum_{|\lambda|\le J} 2^{4|\lambda|}\,\mathbb{E}_a[\exp(\tfrac{6}{\kappa} 2^{|\lambda|} T^{1/2}|E_\lambda|)]^{1/2} \bullet \\
&\qquad \exp(-\tfrac{3}{2\kappa} 2^{|\lambda|} T^{1/2}\kappa_\lambda)\,\mathbb{E}_a[E_\lambda^4]^{1/2} \\
&\lesssim \sum_{|\lambda|\le J} 2^{4|\lambda|} T^{-3/2} T^{-1} 2^{-2|\lambda|} \\
&\lesssim T^{-5/2} 2^{3J} \sim T^{-1}.
\end{aligned}$$

Even more easily, the large deviation estimate bounds the sum over $\mathbb{E}_a[S_3(\lambda)]$:

$$\begin{aligned}
\sum_{|\lambda|\le J} 2^{4|\lambda|}\,\mathbb{E}_a[S_3(\lambda)] &= \sum_{|\lambda|\le J} 2^{4|\lambda|}\,\mathbb{P}_a(|E_\lambda| > \tfrac{\kappa_\lambda}{2}) Z_\lambda^2 \\
&\lesssim \|Q_a a\|_{H^2}^2 \exp(-\tfrac{3}{\kappa} 2^{|\lambda|} T^{1/2}\kappa_\lambda) \\
&\sim T^{-3} \|a\|_{\mathcal{W}^0_{2,2}}.
\end{aligned}$$

The remaining estimates rely on nonlinear approximation theory. Using the characterisation of the Besov space norm by $(s+2)$-regular wavelets (Appendix A.3)

$$\|Q_a a\|^p_{B^{s+2}_{p,\infty}} \sim \sup_{j\geq 0} 2^{jp(s+2+\frac{1}{2}-\frac{1}{p})} \sum_k |Z_{jk}|^p,$$

we infer for all $j \in \mathbb{N}_0$ and $\tau > 0$ by a Chebyshev inequality-type argument the following bound on the cardinality of large wavelet coefficients:

$$|\{k \,|\, |Z_{jk}| \geq \tau\}| \lesssim \|Q_a a\|^p_{B^{s+2}_{p,\infty}} 2^{-jp(s+2+\frac{1}{2}-\frac{1}{p})} \tau^{-p}. \tag{7.1.2}$$

We set $\pi := p \wedge 2$ and obtain the estimate $s + 3(\frac{1}{2} - \frac{1}{\pi}) > 0$ from the restriction $p > \max(\frac{6}{2s+3}, 1)$. The sum involving $S_2(\lambda)$ can be bounded by separate estimates, setting $2^{j_0} \sim T^{\frac{1}{2s+3}}$:

$$\begin{aligned}
\sum_{|\lambda|\leq J} 2^{4|\lambda|}\,\mathbb{E}_a[S_2(\lambda)] &= \sum_{|\lambda|\leq J} 2^{4|\lambda|}\,\mathbb{E}_a[E_\lambda^2]\mathbf{1}_{|Z_\lambda|>\kappa_\lambda/2} \\
&\lesssim \sum_{|\lambda|\leq j_0} 2^{4|\lambda|} T^{-1} 2^{-2|\lambda|} \\
&\quad + \sum_{j>j_0} 2^{4j} T^{-1} 2^{-2j} \|Q_a a\|^\pi_{B^{s+2}_{\pi,\infty}} 2^{-j\pi(s+2+\frac{1}{2}-\frac{1}{\pi})} (\kappa_\lambda/2)^{-\pi} \\
&\lesssim T^{-1} 2^{3j_0} + T^{-1+\frac{\pi}{2}} \|a\|^\pi_{s,\pi,\infty} \sum_{j>j_0} 2^{-j\pi(s+3(\frac{1}{2}-\frac{1}{\pi}))} \\
&\sim T^{-\frac{2s}{2s+3}} + T^{-1+\frac{\pi}{2}} 2^{-j_0\pi(s+3(\frac{1}{2}-\frac{1}{\pi}))} \\
&\sim T^{-\frac{2s}{2s+3}} + T^{\frac{-2s-3+\pi s+\frac{3}{2}\pi-\pi s-\frac{3}{2}\pi+3}{2s+3}} \\
&= 2T^{-\frac{2s}{2s+3}}.
\end{aligned}$$

The same technique, notably estimate (7.1.2), applies to the estimate of the sum over $S_4(\lambda)$, which is deterministic. This time we have to use a more refined analysis by considering the values of Z_λ in certain dyadic intervals. Here finally the logarithmic term enters, because one must choose $2^{j_0} \sim (T(\log T)^{-2})^{\frac{1}{2s+3}}$ for balancing the two appearing sums. We find for $p < 2$

$$\begin{aligned}
\sum_{|\lambda|\leq J} 2^{4|\lambda|}\,\mathbb{E}_a[S_4(\lambda)] &= \sum_{|\lambda|\leq J} 2^{4|\lambda|} Z_\lambda^2 \mathbf{1}_{|Z_\lambda|\leq 2\kappa_\lambda} \\
&\leq \sum_{|\lambda|\leq j_0} 2^{4|\lambda|} 4\kappa_\lambda^2 + \sum_{|\lambda|>j_0} 2^{4|\lambda|} Z_\lambda^2 \sum_{m=0}^\infty \mathbf{1}_{2^{-m}\kappa_\lambda<|Z_\lambda|\leq 2^{-m+1}\kappa_\lambda} \\
&\lesssim 2^{3j_0} T^{-1} (\log T)^2 + \sum_{|\lambda|>j_0} \sum_{m=0}^\infty 2^{4|\lambda|} 2^{-2m+2} \kappa_\lambda^2 \mathbf{1}_{|Z_\lambda|>2^{-m}\kappa_\lambda} \\
&\lesssim \left(\frac{T}{(\log T)^2}\right)^{-\frac{2s}{2s+3}} + \\
&\quad \left(\frac{T}{(\log T)^2}\right)^{-1} \sum_{m=0}^\infty 2^{-2m+2} \sum_{|\lambda|>j_0} 2^{2|\lambda|} \mathbf{1}_{|Z_\lambda|>2^{-m}\kappa_\lambda} \\
&\lesssim \left(\frac{T}{(\log T)^2}\right)^{-\frac{2s}{2s+3}} + \\
&\quad \left(\frac{T}{(\log T)^2}\right)^{-1} \sum_{m=0}^\infty 2^{-2m} \sum_{j>j_0} 2^{2j} 2^{-jp(s+2+\frac{1}{2}-\frac{1}{p})} 2^{mp} \kappa_\lambda^{-p}
\end{aligned}$$

$$\begin{aligned}
&\lesssim \left(\frac{T}{(\log T)^2}\right)^{-\frac{2s}{2s+3}} + \\
&\left(\frac{T}{(\log T)^2}\right)^{-1} 2^{-j_0 p(s+3(\frac{1}{2}-\frac{1}{p})} \left(\frac{T}{(\log T)^2}\right)^{p/2} \sum_{m=0}^{\infty} 2^{-(2-p)m} \\
&\sim \left(\frac{T}{(\log T)^2}\right)^{-\frac{2s}{2s+3}}.
\end{aligned}$$

For the estimate in the case $p \geq 2$ we choose π in the interval $(\frac{6}{2s+3}, 2)$, use the embedding property $B^s_{p,\infty} \subset B^s_{\pi,\infty}$ for $\pi < p$ and repeat the same calculations with p replaced by π.

By putting all estimates together, we find the bound

$$\mathbb{E}_a[\|\hat{y}_T - Q_a a\|^2_{H^2}] \lesssim \left(\frac{T}{(\log T)^2}\right)^{-\frac{2s}{2s+3}},$$

where the constant holds uniformly for $a \in M(s,p,S,\delta)$. □

In the next step we construct an operator $\widehat{Q}_T$ from the observations up to time T, which is close to the true covariance operator. We could, of course, use the results for Q_T from Proposition 7, but it is even simpler to use the relationship $q'_a(t) = Q_a a(-t)$ for $t \in (0, r]$ which is deduced from (2.3.15):

$$q'_a(t) = \int_{-r}^0 q_a(t+s)\, da(s) = \int_{-r}^0 q_a(-t-s)\, da(s) = Q_a a(-t).$$

Writing $q_a(t) = q_a(0) + \int_0^t q'_a(u)du$, we can thus determine q_a from the knowledge of $q_a(0)$ and $Q_a a$ and derive an estimator from estimators for these two quantities. This is exactly the construction method of $\widehat{Q}_T$ we shall adopt. We thus avoid calculating and smoothing the function q_T (only $q_T(0,0)$ is needed).

Theorem 7. *Let $s \in (0, s_{max}]$, $S > 0$, $\max(\frac{6}{2s+3}, 1) < p < \infty$ and $\delta > 0$ be given. Introduce the integral operator $\widehat{Q}_T$ with convolution kernel*

$$\hat{q}_T(u) := \frac{1}{T}\int_0^T X(t)^2\, dt + \int_{-|u|}^0 \hat{y}_T(v)\, dv, \quad u \in [-r, r],$$

i.e. $\widehat{Q}_T\mu(t) := \int_{-r}^0 \hat{q}_T(t-u)\, d\mu(u)$ for $t \in [-r, 0]$, $\mu \in M([-r, 0])$. Then the operator $\widehat{Q}_T : \mathcal{W}^0_{2,2} \to H^2([-r, 0])$ is continuous. It is invertible on the set

$$\mathcal{C}_T := \{\|Q_a^{-1}\|_{H^2 \to \mathcal{W}^0_{2,2}} \|Q_a - \widehat{Q}_T\|_{\mathcal{W}^0_{2,2} \to H^2} \leq \tfrac{1}{2}\}.$$

Define the $\sigma(X(t), -r \leq t \leq T)$-measurable estimator $\hat{a}_T$ by

$$\hat{a}_T := \begin{cases} \min\left(S\|\widehat{Q}_T^{-1}\hat{y}_T\|^{-1}_{0,2,2}, 1\right)\widehat{Q}_T^{-1}\hat{y}_T, & \text{if } \widehat{Q}_T \text{ is invertible,} \\ 0, & \text{otherwise.} \end{cases}$$

Then the following asymptotic upper bound holds for $T \to \infty$:

$$\sup_{a \in M(s,p,S,\delta)} \mathbb{E}_a[\|\hat{a}_T - a\|^2_{\mathcal{W}^0_{2,2}}] \lesssim \left(\frac{T}{(\log T)^2}\right)^{-\frac{2s}{2s+3}}.$$

Proof. Due to $\hat{y}_T \in H^2([-r,0])$ the kernel $\hat{q}_T|_{[0,r]}$ is an element of $H^3([0,r])$ and the continuity of $\widehat{Q}_T : \mathcal{W}^0_{2,2} \to H^2([-r,0])$ follows from Lemma 5 and estimate (3.3.2). Formally, the Neumann series expansion yields for $\widehat{Q}_T^{-1}$

$$\widehat{Q}_T^{-1} = (\mathrm{Id} - Q_a^{-1}(Q_a - \widehat{Q}_T))^{-1} Q_a^{-1} = \sum_{m=0}^{\infty} (Q_a^{-1}(Q_a - \widehat{Q}_T))^m Q_a^{-1}.$$

This expansion is as usually a posteriori justified for the operators $\widehat{Q}_T, Q_a : \mathcal{W}^0_{2,2} \to H^2([-r,0])$ if $\|Q_a^{-1}\|_{H^2 \to \mathcal{W}^0_{2,2}} \|Q_a - \widehat{Q}_T\|_{\mathcal{W}^0_{2,2} \to H^2} < 1$ holds:

$$\begin{aligned}
\|\widehat{Q}_T^{-1}\|_{H^2 \to \mathcal{W}^0_{2,2}} &= \left\| \sum_{m=0}^{\infty} (Q_a^{-1}(Q_a - \widehat{Q}_T))^m Q_a^{-1} \right\|_{H^2 \to \mathcal{W}^0_{2,2}} \\
&\leq \sum_{m=0}^{\infty} \|Q_a^{-1}\|^m_{H^2 \to \mathcal{W}^0_{2,2}} \|Q_a - \widehat{Q}_T\|^m_{\mathcal{W}^0_{2,2} \to H^2} \|Q_a^{-1}\|_{H^2 \to \mathcal{W}^0_{2,2}} \\
&= \frac{\|Q_a^{-1}\|_{H^2 \to \mathcal{W}^0_{2,2}}}{1 - \|Q_a^{-1}\|_{H^2 \to \mathcal{W}^0_{2,2}} \|Q_a - \widehat{Q}_T\|_{\mathcal{W}^0_{2,2} \to H^2}} < \infty.
\end{aligned}$$

Hence, on $\mathcal{C}_T$ the operator $\widehat{Q}_T$ is invertible with

$$\|\widehat{Q}_T^{-1}\|_{H^2 \to \mathcal{W}^0_{2,2}} \leq 2\|Q_a^{-1}\|_{H^2 \to \mathcal{W}^0_{2,2}}.$$

In order to bound the probability of $\mathcal{C}_T$ from below, we use the estimate $\|Q_a - \widehat{Q}_T\|_{\mathcal{W}^0_{2,2} \to H^2} \leq C_1 \|q_a - \hat{q}_T\|_{H^2([-r,0])}$, $C_1 > 0$ some universal constant, derived from Lemma 5 with $k(t) = (q_a - \hat{q}_T)(-t)$ and from relation (3.3.2) (on $\mathrm{span}(\delta_{-r}, \delta_0)$ this is immediate). From Proposition 13 we know

$$\mathbb{E}_a[\|q_a' - \hat{q}_T'\|^2_{H^2}] = \mathbb{E}_a[\|Q_a a(-\bullet) - \hat{y}_T(-\bullet)\|^2_{H^2}] \lesssim \left(\frac{T}{(\log T)^2} \right)^{-\frac{2s}{2s+3}}.$$

Furthermore, from Proposition 7 with $\alpha = 0$ and $p = 2$ follows

$$\mathbb{E}_a \left[\left| q_a(0) - \frac{1}{T} \int_0^T X(u)^2 \, du \right|^2 \right] \lesssim T^{-1}.$$

We conclude

$$\mathbb{E}_a[\|q_a - \hat{q}_T\|^2_{H^3}] \lesssim \left(\frac{T}{(\log T)^2} \right)^{-\frac{2s}{2s+3}}. \tag{7.1.3}$$

In all the estimates the constants can be chosen to hold for all $a \in M(s,p,S,\delta)$ by the uniformity established in Propositions 13 and 7. Finally, Chebyshev's inequality yields with constants $C_2, C_3 > 0$

$$\begin{aligned}
\sup_{a \in M(s,p,S,\delta)} \mathbb{P}_a(\Omega \setminus \mathcal{C}_T) &\leq \sup_{a \in M(s,p,S,\delta)} \mathbb{P}_a(\|q_a - \hat{q}_T\|_{H^2} > C_2) \\
&\leq \sup_{a \in M(s,p,S,\delta)} \mathbb{E}_a[\|q_a - \hat{q}_T\|^2_{H^2}] C_2^{-2} \\
&\leq C_3 \left(\frac{T}{(\log T)^2} \right)^{-\frac{2s}{2s+3}}.
\end{aligned}$$

It therefore suffices to work on the set $\mathcal{C}_T$, because on its complement the loss is bounded by $2S$ and hence the squared risk tends to zero uniformly with rate $(T/(\log T)^2)^{-\frac{2s}{2s+3}}$. On $\mathcal{C}_T$ we obtain

$$\begin{aligned}
\|\hat{a}_T - a\|_{0,2,2} &= \|\widehat{Q}_T^{-1}\hat{y}_T - Q_a^{-1}Q_a a\|_{0,2,2} \\
&\le \|\widehat{Q}_T^{-1}\|_{H^2\to\mathcal{W}^2_{0,2,2}}\|\hat{y}_T - Q_a a\|_{H^2} + \|\widehat{Q}_T^{-1} - Q_a^{-1}\|_{H^2\to\mathcal{W}^2_{0,2,2}}\|Q_a a\|_{H^2} \\
&\le 2\|Q_a^{-1}\|_{H^2\to\mathcal{W}^2_{0,2,2}}\|\hat{y}_T - Q_a a\|_{H^2} \\
&\quad + \|\widehat{Q}_T^{-1}\|_{H^2\to\mathcal{W}^2_{0,2,2}}\|Q_a - \widehat{Q}_T\|_{\mathcal{W}^2_{0,2,2}\to H^2}\|Q_a^{-1}\|_{H^2\to\mathcal{W}^2_{0,2,2}}\|Q_a a\|_{H^2} \\
&\le 2\|Q_a^{-1}\|_{H^2\to\mathcal{W}^2_{0,2,2}}\|\hat{y}_T - Q_a a\|_{H^2} \\
&\quad + 2\|Q_a^{-1}\|^2_{H^2\to\mathcal{W}^2_{0,2,2}}\|Q_a - \widehat{Q}_T\|_{\mathcal{W}^2_{0,2,2}\to H^2}\|Q_a a\|_{H^2} \\
&\lesssim \|\hat{y}_T - Q_a a\|_{H^2} + \|q_a - \hat{q}_T\|_{H^2}.
\end{aligned}$$

By Corollary 7 the last estimate holds with a uniform constant for all $a \in M(s,p,S,\delta)$.

From the preceding Proposition 13 and the estimate (7.1.3) we conclude

$$\sup_{a\in M(s,p,S,\delta)} \mathbb{E}_a[\|\hat{a}_T - a\|^2_{0,2,2}\mathbf{1}_{\mathcal{C}_T}] \lesssim T^{-\frac{2s}{2s+3}},$$

which accomplishes the proof of the asymptotic risk upper bound. □

Remarks 12.

- *One might want to consider the submodel in which the weights do not include any point measures, i.e. the weight space $B^s_{p,\infty}$ instead of $\mathcal{W}^s_{p,\infty}$. For this one can project the estimator $\hat{a}_T$ onto $L^2([-r,0])$ by neglecting the point measure part. Another possibility is to project $\hat{y}_T$ H^2-orthogonally to $\widehat{Q}_T(L^2)$. In both cases the asymptotic risk rate does not increase, because under the submodel assumption the errors due to these corrections are at most of the order of the risk.*

- *Our method differs from the classical wavelet thresholding algorithm for density estimation or regression due to the ill-posedness involved. Our threshold κ_λ depends on the resolution level $|\lambda|$, because the intensity of the noise coefficients measured in H^2-norm is of order $2^{2|\lambda|}\,\mathbb{E}[E_\lambda^2]^{1/2} \sim 2^{|\lambda|}T^{-1/2}$. Furthermore, we only need to calculate the wavelet coefficients up to a resolution level $2^J \sim T^{1/2}$ and it is not necessary to suppose that the weight also has some regularity with respect to an L^2-Sobolev space. The reason for these two phenomena is that the condition $p > \max(\frac{6}{2s+3}, 1)$ is so strong that the linear approximation error is already sufficiently small. That the restriction $p > \frac{6}{2s+3}$ or equivalently $s - 3(\frac{1}{p} - \frac{1}{2}) > 0$ is really necessary will follow from the lower bound in the next section.*

- *Another deviation from classical results is the factor $T(\log T)^{-2}$, which stems from the fact that we only have finite exponential moments of the noise and not higher Gaussian-like moments, for which the factor $T(\log T)^{-1}$ is achieved. Since q_T and b_T are quadratic in X, this can probably not be improved on. It might however be possible to employ some uniform functional central limit theorem and apply results similar to those of Neumann and von Sachs (1995) in order to enlarge the factor to $T(\log T)^{-1}$.*

- *Concerning implementation issues it should be mentioned that it seems rather difficult to determine $\alpha(S,\delta)$ and $\beta(S,\delta)$. Numerically challenging is also the inversion algorithm for the calculation of $\widehat{Q}_T^{-1}\hat{y}_T$. The investigation of this*

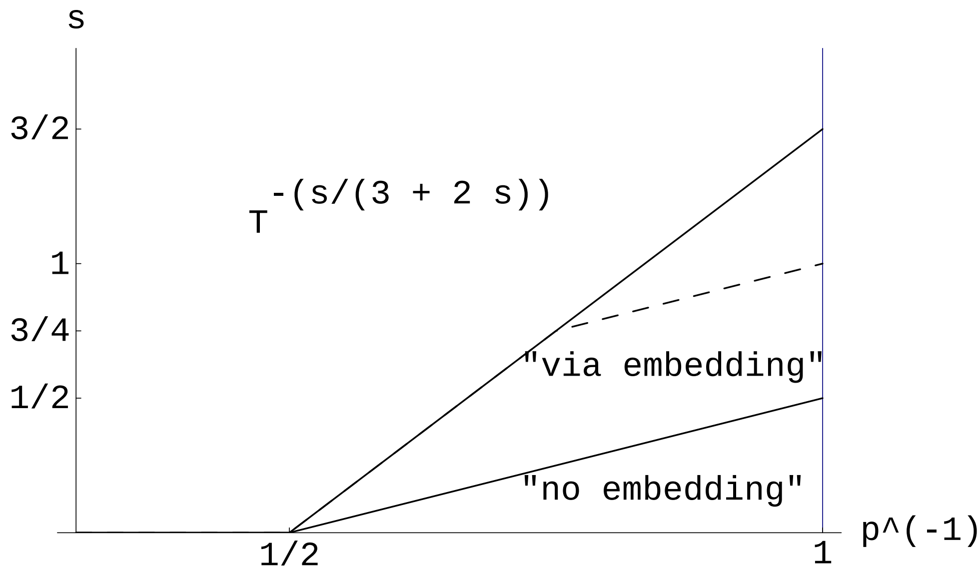

Figure 7.1.1: Adaptive L^2-risk for weights in $\mathcal{W}^s_{p,\infty}$. The dashed line corresponds to the embedding of a piecewise constant weight function.

problem is ongoing work with Albert Cohen and Marc Hoffmann, Paris. Since we have got estimation errors anyway, we can employ an iterative method, which in each step combines a thresholding algorithm with a step from an iterative solver for linear systems, for instance a gradient method. This algorithm seems to yield rate-optimal results with a small amount of computing time and storage. The approach here is closely related to the wavelet-vaguelette decomposition method proposed by Donoho (1995). Due to boundary effects it is however difficult to determine whether the functions $\widehat{Q}_T^{-1}\psi_\lambda$, when properly scaled, fulfil the vaguelette properties.

- *Having obtained results for wavelet thresholding in $H^2([-r,0])$-loss, it is not difficult to transfer the classical theory for L^p-loss to L^p-Sobolev loss and hence to bound the risk of the thresholding estimator of the weight a with respect to an L^p-type loss function. The abstract approach has been investigated by Kerkyacharian and Picard (2000). For weights in $\mathcal{W}^s_{\pi,\infty}$ the L^p-risk rate $T^{-\frac{s}{2s+3}}$ is obtained, provided $\pi > \max(\frac{3p}{3+ps}, 1)$ is satisfied. Other thresholding methods than hard thresholding or also adaptive kernel methods could also be applied (cf. Härdle et al. (1998) and the references given there).*

Example 7. *Which rate of convergence do we obtain for the L^2-risk of the weight function $g = -\mathbf{1}_{[-\frac{1}{2},0]}$ in the case $r = 1$? Since g lies in the spaces H^s for all $s < \frac{1}{2}$ (Appendix A.2) we obtain roughly the rate $T^{-1/8}$ for the Galerkin estimator. With respect to the scale of Besov spaces g lies in $B^s_{1/s,\infty}$, $s \in (0,1)$ (Appendix A.2). For the risk estimate of the wavelet thresholding estimator we need the restriction $p = \frac{1}{s} > \frac{6}{3+2s}$ such that the maximal value of s, we can use for the rate, is $s \approx \frac{3}{4}$, as shown in Figure 7.1.1. The adaptive estimation rate is therefore approximately $T^{-1/6}$. This is slow, but significantly better than the non-adaptive rate. Similarly, smooth weight functions only having a jump in the first derivative at one point belong to H^s for $s < \frac{3}{2}$ and to $B^2_{1,\infty}$, whence the Galerkin estimator converges roughly with the rate $T^{-1/4}$ and the wavelet thresholding estimator roughly with the rate $T^{-2/7}$.*

If the estimation theory also worked for $p \leq 1$*, we would even obtain the rate* $T^{-3/10}$ $(s = \frac{9}{4},\ p = \frac{4}{5})$*. Note that the more regular the weight functions are, the less the convergence rates differ in the adaptive and in the non-adaptive case.*

7.2 Lower risk bound: sparse case

The content of this section is to show that the adaptive wavelet thresholding estimator is – up to logarithmic factors – rate-optimal with respect to the L^2-risk function, in the sense that one cannot improve on the restriction $p > \frac{6}{3+2s}$ in order to obtain the speed of convergence $T^{-\frac{s}{2s+3}}$ for weights in $\mathcal{W}^s_{p,\infty}$. For smaller values of p the rate of convergence is indeed worse and is obtained by embedding $\mathcal{W}^s_{p,\infty}$ to $\mathcal{W}^\sigma_{\pi,\infty}$ with some properly chosen $\sigma < s$ and $\pi > p$, as it was done in Example 7. In the sequel, we merely have to assume $s + \frac{1}{2} - \frac{1}{p} > 0$ in order to have the embedding $\mathcal{W}^s_{p,\infty} \subset \mathcal{W}^0_{2,2}$ and a well-defined risk.

The expression"sparse case" for the estimation of functions with L^p-regularity for $p < 2$ is explained by the fact that the most difficult functions to estimate are those which have very localized irregular peaks (cf. the discussion in Härdle et al. (1998, Section 10.4)).

Theorem 8. *Let* $s > 0$*,* $p > 0$*,* $S > 0$ *and* $\delta > 0$ *be given with* $s + \frac{1}{2} - \frac{1}{p} > 0$*, i.e.* $p > \frac{2}{2s+1}$*, and such that* $M(s,p,S,\delta)$ *has nonempty interior in* $\mathcal{W}^s_{p,\infty}$*. Then the following asymptotic minimax lower bound holds for* $T \to \infty$*:*

$$\inf_{\hat{a}_T} \sup_{a \in M(s,p,S,\delta)} \mathbb{E}_a[\|\hat{a}_T - a\|^2_{\mathcal{W}^0_{2,2}}]^{1/2} \gtrsim \left(\frac{T}{\log T}\right)^{-\frac{s+\frac{1}{2}-\frac{1}{p}}{2s+3-\frac{2}{p}}},$$

where the infimum is taken over all $\mathcal{F}^X_T$*-measurable estimators* $\hat{a}_T$*.*

Proof. As in Chapter 6 we build from a weight a_0 in the interior of $M(s,p,S,\delta)$ a family of local alternatives (a_k). Choose a compactly supported s-regular wavelet basis in $L^2(\mathbb{R})$ and denote by R_j a maximal set of integers with $\operatorname{supp}(\psi_{jk}) \subset [-r, 0]$ and $\operatorname{supp}(\psi_{jk}) \cap \operatorname{supp}_(\psi_{jk'}) = \varnothing$ for all $k, k' \in R_j$, $k \neq k'$. For any $k \in R_j$ we set

$$a_{jk} := a_0 + \gamma \psi_{jk}$$

with $\gamma = \gamma(T) \sim 2^{-j(T)(s+\frac{1}{2}-\frac{1}{p})}$ such that $\|a_{jk}\|_{s,p,\infty} \leq S$ and $v_0(a_{jk}) \leq -\delta$ are satisfied, hence $a_{jk} \in B(s,S,p,\delta)$ holds true. We briefly interrupt the proof for stating a classical lemma for the lower bound proofs in the sparse case.

Lemma 15. *Suppose the likelihood ratio satisfies*

$$\mathbb{P}_{a_{jk}}\Big(\log(\Lambda_T(X^{(a_0)}, X^{(a_{jk})})) \geq -j\Big) \geq \pi_0 > 0$$

uniformly for all a_{jk}*. Then for* $\mathcal{F}^X_T$*-measurable estimators* $\hat{a}_T$ *the following lower bound holds:*

$$\inf_{\hat{a}_T} \sup_{a \in M(s,p,S,\delta)} \mathbb{E}_a[\|\hat{a}_T - a\|^2_{L^2}]^{1/2} \gtrsim \gamma(T).$$

Proof of the Lemma. This is an adapted version of Härdle et al. (1998, Lemma 10.1). Only choose $v^k_T = |R_j| \sim 2^j$, substituting n by T. Then the lower bound for the risk follows easily from the estimate

$$\liminf_{T\to\infty} \inf_{\hat{a}_T} \sup_{a \in M(s,p,S,\delta)} \mathbb{P}_a(\|\hat{a}_T - a\|_{L^2} \geq \gamma(T)) > 0$$

on the set $\{\|\hat{a}_T - a\|_{L^2} \geq \gamma(T)\}$ and hence everywhere. □

Let us resume the proof of the theorem. Exactly as in the proof of the lower bound in Theorem 5, we obtain

$$\mathbb{E}_{a_{jk}}[\log(\Lambda_T(X^{(a_0)}, X^{(a_{jk})})^2] \lesssim \gamma^4 T^2 2^{-4j} + \gamma^4 T 2^{-2j}$$

with a uniform constant for all a_{jk}. Thus, by Chebyshev's inequality the requirements of Lemma 15 are satisfied, when we balance the conditions on γ by choosing $2^{(2s+3-\frac{2}{p})j} \sim T/\log(T)$ such that

$$\gamma^4 T^2 2^{-4j} + \gamma^4 T 2^{-2j} \sim \left(\frac{T}{\log(T)}\right)^{\frac{-4(s+\frac{1}{2}-\frac{1}{p})-4}{2s+3-\frac{2}{p}}} T^2 \sim (\log T)^2 \sim j^2$$

holds. Thus, the lower bound follows. □

We have obtained a fairly complete picture of the minimax rates for the L^2-risk of certain Besov regularity classes $M(s,p,S,\delta)$. This is the content of the next corollary and presented graphically in the (s,p^{-1})-plane in Figure 7.1.1.

Corollary 11. *Assume that $s>0$, $1<p<\infty$, $S>0$ and $\delta>0$ are given such that $M(s,p,S,\delta)$ has nonempty interior in $\mathcal{W}^s_{p,\infty}$. In what follows the infima are taken over all $\mathcal{F}^X_T$-measurable estimators $\hat{a}_T$.*

1. *If*

$$\inf_{\hat{a}_T} \sup_{a\in M(s,p,S,\delta)} \mathbb{E}_a[\|\hat{a}_T - a\|^2_{\mathcal{W}^0_{2,2}}]^{1/2} \gtrsim \left(\frac{T}{\log T}\right)^{-\frac{s}{2s+3}}$$

   holds, then $p \geq \frac{6}{2s+3}$ follows.

2. *In the case $p > \frac{6}{2s+3}$, i.e. $s + 3(\frac{1}{2}-\frac{1}{p}) > 0$, we have*

$$\left(\frac{T}{\log T}\right)^{-\frac{s}{2s+3}} \lesssim \inf_{\hat{a}_T} \sup_{a\in M(s,p,S,\delta)} \mathbb{E}_a[\|\hat{a}_T - a\|^2_{\mathcal{W}^0_{2,2}}]^{1/2} \lesssim \left(\frac{T}{(\log T)^2}\right)^{-\frac{s}{2s+3}}.$$

3. *For $p \leq \frac{6}{2s+3}$ we obtain*

$$\left(\frac{T}{\log T}\right)^{-\frac{s+\frac{1}{2}-\frac{1}{p}}{2s+3-\frac{2}{p}}} \lesssim \inf_{\hat{a}_T} \sup_{a\in M(s,p,S,\delta)} \mathbb{E}_a[\|\hat{a}_T - a\|^2_{\mathcal{W}^0_{2,2}}]^{1/2} \lesssim T^{-\frac{\sigma+\frac{1}{2}-\frac{1}{p}}{2\sigma+3-\frac{2}{p}}}$$

   for all $\sigma < s$.

Proof.

1. This follows from Theorem 8, using

$$\frac{s+\frac{1}{2}-\frac{1}{p}}{2s+3-\frac{2}{p}} \geq \frac{s}{2s+3} \iff s+\tfrac{3}{2}-\tfrac{3}{p} \geq 0.$$

2. The upper bound is due to Theorem 7 and the lower bound is due to Theorem 5, taking into account the Remarks 10 concerning $B^s_{\infty,\infty}$.

3. The lower bound is just Theorem 8. For the upper bound use the embbedding $\mathcal{W}^s_{p,\infty} \subset \mathcal{W}^\tau_{\pi,\infty}$ for all $\tau < \frac{3}{2}(s+\frac{1}{2}-\frac{1}{p})$ and

$$\pi := (\tau - s + \tfrac{1}{p})^{-1} > \tfrac{6}{2\tau+3}.$$

Due to $\pi > p > 1$ we can apply Theorem 7 for $B(\tau, \pi, S', \delta)$, S' chosen appropriately, and obtain for all $\sigma < \frac{2}{3}\tau - \frac{1}{2} + \frac{1}{p}$ the rate

$$\left(\frac{T}{(\log T)^2}\right)^{-\frac{\tau}{2\tau+3}} \leq T^{-\frac{\sigma+\frac{1}{2}-\frac{1}{p}}{2\sigma+3-\frac{2}{p}}}$$

and thus also for all $\sigma < s$.

□

Chapter 8

Hypothesis testing

Often, statistical methods are not applied for estimation purposes, but to confirm or to reject a certain hypothesis. The knowledge of the mean square error in an estimation procedure yields a confidence interval, but this construction is in general not optimal. The motivation for this chapter is the development of a testing procedure that decides whether an observed trajectory of an affine SDDE has some memory effect or is just a simple Ornstein-Uhlenbeck process. As it turns out, the test will also apply to the more general problem to decide whether a_0 from the generalized L^2-space $\mathcal{W}^0_{2,2}$ is the true underlying weight. The nonparametric alternative will consist of those weights in this class that lie outside an $\mathcal{W}^0_{2,2}$-ball around a_0 of radius ρ and satisfy a Sobolev regularity condition of order s.

A test is constructed that can decide between hypothesis and alternative with small error probabilities if the separation rate satisfies $\rho \gtrsim T^{-\frac{s}{2s+2.5}}$ for $T \to \infty$. Note that the test derived from the Galerkin estimator would by Chebyshev's inequality separate hypothesis and alternative only for $\rho \gtrsim T^{-\frac{s}{2s+3}}$. This phenomenon of rate improvement is familiar in nonparametric testing theory for L^2-spaces. The ideas and proofs are derived from a related Gaussian problem, which will be investigated as a model problem in parallel with the original SDDE testing problem. The first section states the testing problem thoroughly and introduces the test statistics used. In the second section the minimal separation rate between hypothesis and alternative for which the test works is determined. In the last section, this rate is shown to be optimal among all conceivable tests.

8.1 Construction of a test

We take advantage of the mapping properties of the covariance operator (Theorem 2) and use the space $\mathcal{W}^0_{2,2}$ as the parameter space for the weights. A thorough discussion of testing a single hypothesis versus a nonparametric set of alternatives in the signal plus white noise model is given in the series of papers Ingster (1993a; Ingster (1993b; Ingster (1993c). Classically, the set of alternatives $\mathcal{K}_\rho$ is a compact Sobolev ball where an L^2-ball of radius $\rho > 0$ around the hypothesis is cut out. We think of having fixed the error level of the first kind $\alpha \in (0, 1)$ and of the second kind $\beta \in (0, 1)$ and ask for the smallest separation value ρ in the alternative such that these error bounds are not transgressed. This value of ρ will be determined asymptotically with respect to the observation time T. Having fixed the general ideas we now define our testing problem precisely.

Definition 12. *Let $s>0$, $S>\rho>0$ and $\delta>0$ be given. For $a_0\in\mathcal{W}^s_{2,2}$ with $v_0(a_0)<-\delta$ set*

$$\mathcal{H}_0 := \{a_0\},$$
$$\mathcal{K}_\rho := \{a\in\mathcal{W}^s_{2,2} \mid \|a-a_0\|_{s,2,2}\le S,\ v_0(a)\le-\delta,\ \|a-a_0\|_{0,2,2}\ge\rho\}.$$

The set $\mathcal{H}_0$ encodes the hypothesis *that the true underlying weight is a_0 and the set $\mathcal{K}_\rho$ corresponds to the* alternative *that one of its elements is the true underlying weight.*

The subsequent results remain true if it is only assumed that a_0 is in $\mathcal{W}^0_{2,2}$ and if one considers the alternative

$$\mathcal{K}_\rho := \{a\in\mathcal{W}^0_{2,2} \mid \|Q_a(a-a_0)\|_{s+2,2,2}\le S,\ v_0(a)\le-\delta,\ \|a-a_0\|_{0,2,2}\ge\rho\}.$$

Drawing a parallel with a signal plus noise model from the experience of the estimation problem, we might view b_T as noisy observation of Q_Ta under $\mathbb{P}_a$. Heuristically, the difference $b_T-Q_Ta_0$ should be a comparatively small function under $\mathbb{P}_{a_0}$, but contain a significant signal under $\mathbb{P}_a$ for weights a in some distance of a_0. Observe, however, the difficulties arising through the heteroskedastic nature of the noise and through the ill-posedness induced by Q_T. The corresponding Gaussian shift experiment serves as a good intuition in order to understand the subsequent construction of the test statistic and the proofs of its asymptotic behaviour. We therefore include the calculations for this model problem.

▷ **Model problem.** Suppose Q is any fixed covariance operator from Definition 4 with a weight from $\mathcal{W}^s_{2,2}$. Let $\Gamma\sim N(0,Q)$ be a centred Gaussian $C([-r,0])$-valued random variable. Based on the observation

$$Y := Qa + T^{-1/2}\Gamma,\quad a\in\mathcal{H}_0\cup\mathcal{K}_\rho,$$

we want to decide whether a lies in $\mathcal{H}_0$ or $\mathcal{K}_\rho$. We consider the fastest convergence rate in $\rho(T)\to0$ for $T\to\infty$, for which this can still be accomplished. By applying $Q^{-1/2}$ to Y we get a white noise model with signal $Q^{1/2}a$ and it only remains to take care of the fact that the signals corresponding to hypothesis and alternative are separated in the norm $\|Q^{1/2}\bullet\|_{L^2}$, i.e. roughly in the weaker H^{-1}-norm. This approach already indicates that we shall end up with a rate $T^{-\frac{s}{2(s+1)+0.5}}$ instead of classically $T^{-\frac{s}{2s+0.5}}$.

When using the correct functional norm, namely the $H^2([-r,0])$-norm, we can apply the test statistic known for the white noise model. For the sake of simplicity we shall explicitly use wavelet bases. They will only be used for smoothing the data (and not for inversion) so that finite element or kernel methods should be working in the same way.

Definition 13. *Given an $(s\vee2)$-regular wavelet basis (ψ_λ) on $L^2([-r,0])$, the quantities b_T and Q_T from Definition 6 and the weight a_0 from Definition 12, we define the test statistic*

$$\mathcal{T}_{T,J} := \frac{1}{T}\sum_{|\lambda|\le J} 2^{4|\lambda|}\langle b_T-Q_Ta_0,\psi_\lambda\rangle^2$$

for $T>0$, $J\in\mathbb{N}$. We choose $2^{J(T)}\sim T^{\frac{1}{2s+2.5}}$, set $\mathcal{T}:=\mathcal{T}_{T,J(T)}$ and suppress the time dependence of $\mathcal{T}$. For a given $\alpha\in(0,1)$ we employ the following decision function *φ:*

$$\varphi := \mathbf{1}_{\{\mathcal{Z}>q_{1-\alpha}\}}\ \textit{with}\ \mathcal{Z} := \frac{\mathcal{T}-\mathbb{E}_{a_0}[\mathcal{T}]}{\operatorname{Var}_{a_0}[\mathcal{T}]^{1/2}},$$

where $q_{1-\alpha}$ *denotes the* $(1-\alpha)$*-quantile of* $\mathcal{Z}$ *under* $\mathbb{P}_{a_0}$*. The non-randomized decision is taken according to the rule that the hypothesis is accepted for* $\varphi = 0$ *and it is rejected in favour of the alternative for* $\varphi = 1$*.*

Remarks 13.

- *From the wavelet characterisation of* $H^2([-r, 0])$ *we obtain*
$$\mathcal{T}_{T,j} \sim \tfrac{1}{T}\|P_J(b_T - Q_T a_0)\|_{H^2}^2.$$
- *By following this decision rule we automatically bound the error of the first kind:* $\mathbb{P}_{a_0}(\varphi = 1) = \alpha$*. Observe that by Chebyshev's inequality* $q_{1-\alpha} \leq \alpha^{-1/2}$ *holds. In practice, similar moment approximations or Monte Carlo-simulations may be used in order to determine* $q_{1-\alpha}$*, because it seems rather unlikely that the exact law of* $\mathcal{T}$ *can be determined. For fixed* J *the central limit theorem for mixing processes could be applied as* $T \to \infty$*. In our case, however,* J *tends to infinity with* $T \to \infty$ *so that a very strong infinite-dimensional central limit theorem would be needed.*
- *Compared with the calculation of the Galerkin estimator, the test statistic* $\mathcal{T}$ *is even easier to determine, because we need not solve a multidimensional linear system. Basically, we compare* $Q_{a_0}a_0$ *and* $Q_a a$ *for* $a \in \mathcal{K}_\rho$ *so that the whole analysis takes place in the range of the covariance operators, which is exactly* $H^2([-r, 0])$ *for the domain* $\mathcal{W}_{2,2}^0$*.*

The test can easily be adapted to our model problem. In fact, originally it was derived from it.

▷ **Model problem.** We only substitute the corresponding expressions in $\mathcal{T}$ and take care of the scaling factor T in Q_T and b_T:
$$\mathcal{T}_{T,J} = T \sum_{|\lambda| \leq J} 2^{4|\lambda|}\langle Y - Qa_0, \psi_\lambda\rangle^2.$$
Then we use the same definition of $\mathcal{Z}$ and of the decision function φ.

8.2 Test asymptotics

The whole argument to find the upper bound relies on asymptotic estimates for the expected value and the variance of $\mathcal{T}$ under the different laws corresponding to $\mathcal{H}_0$ and $\mathcal{K}_\rho$. The gain in the convergence rate compared to the estimation problem is due to the fact that under $\mathbb{P}_{a_0}$ the standard deviation of $\mathcal{T}$ is asymptotically significantly smaller than its expected value, which on the other hand demands much more subtle estimates in the proofs.

Proposition 14. *For the moments of* $\mathcal{T}$ *under* $\mathbb{P}_{a_0}$ *we obtain*
$$\mathbb{E}_{a_0}[\mathcal{T}] = \sum_{|\lambda| \leq J(T)} 2^{4|\lambda|}\langle Q_{a_0}\psi_\lambda, \psi_\lambda\rangle, \qquad \mathrm{Var}_{a_0}[\mathcal{T}] \lesssim 2^{5J(T)}.$$

Proof. By the Fubini theorem for stochastic integrals Protter (1992, Thm. 46) and the Itô isometry the expected value of $\mathcal{T}$ is easily determined:
$$\begin{aligned}
\mathbb{E}_{a_0}[\mathcal{T}] &= \frac{1}{T}\sum_{|\lambda| \leq J(T)} 2^{4|\lambda|}\,\mathbb{E}_{a_0}[\langle b_T - Q_T a_0, \psi_\lambda\rangle^2] \\
&= \frac{1}{T}\sum_{|\lambda| \leq J(T)} 2^{4|\lambda|}\,\mathbb{E}_{a_0}\left[\left(\int_0^T \langle X(t+\bullet), \psi_\lambda\rangle\, dW(t)\right)^2\right] \\
&= \sum_{|\lambda| \leq J(T)} 2^{4|\lambda|}\langle Q_{a_0}\psi_\lambda, \psi_\lambda\rangle.
\end{aligned}$$

▷ **Model problem.** Since $T^{-1/2}(b_T - Q_T a_0)$ has the same covariance structure under $\mathbb{P}_{a_0}$ as Γ, the expected value equals $\sum_{|\lambda|\leq J(T)} 2^{4|\lambda|}\langle Q\psi_\lambda, \psi_\lambda\rangle$. For the variance term we use the characterisation of the H^2-norm by 2-regular wavelets (Appendix A.3) and the mapping properties of Q (Theorem 5):

$$\begin{aligned}
\mathrm{Var}_{a_0}[\mathcal{T}] &= \sum_{|\lambda|,|\lambda'|\leq J(T)} 2^{4(|\lambda|+|\lambda'|)}\,\mathrm{Cov}[\langle\Gamma,\psi_\lambda\rangle^2, \langle\Gamma,\psi_{\lambda'}\rangle^2] \\
&= \sum_{|\lambda|,|\lambda'|\leq J(T)} 2^{4(|\lambda|+|\lambda'|)} 2\langle Q\psi_\lambda, \psi_{\lambda'}\rangle^2 \\
&\lesssim \Big\| \sum_{|\lambda|\leq J(T)} 2^{2|\lambda|} Q\psi_\lambda \Big\|_{H^2}^2 \\
&\sim \Big\| \sum_{|\lambda|\leq J(T)} 2^{2|\lambda|} \psi_\lambda \Big\|_{L^2}^2 \\
&= \sum_{|\lambda|\leq J(T)} 2^{4|\lambda|} \\
&\sim 2^{5J(T)}.
\end{aligned}$$

The estimation of higher moments in the SDDE case is more complicated, because we lose too much by a direct application of the Cauchy-Schwarz inequality. For λ, λ' with $|\lambda|, |\lambda'| \leq J$ we aim at estimating the covariance term

$$C_{\lambda,\lambda'} := \mathrm{Cov}_{a_0}[\langle b_T - Q_T a_0, \psi_\lambda\rangle^2,\ \langle b_T - Q_T a_0, \psi_{\lambda'}\rangle^2].$$

Let us introduce the processes $A(t) := \langle X(t+\bullet), \psi_\lambda\rangle$ and $B(t) := \langle X(t+\bullet), \psi_\lambda\rangle$, $t \geq 0$, which are stationary Gaussian processes under $\mathbb{P}_a$ with $A(t) \sim N(0, \langle Q_a\psi_\lambda, \psi_\lambda\rangle)$ and $B(t) \sim N(0, \langle Q_a\psi_{\lambda'}, \psi_{\lambda'}\rangle)$. In the sequel the estimate with $m \in \mathbb{N}$, $0 \leq a < b$

$$\mathbb{E}_a\left[\left(\int_a^b A(s)\,dW(s)\right)^{2m}\right] \lesssim (b-a)^m \langle Q_a\psi_\lambda, \psi_\lambda\rangle^m \lesssim (b-a)^m 2^{-2m|\lambda|}, \quad (8.2.1)$$

and its analogue for $B(t)$ in terms of λ' will be needed. They are derived from the martingale moment inequality Karatzas and Shreve (1991, Prop. 3.26) and Corollary 13. We obtain the following decomposition by the Fubini theorem for stochastic integrals and the Itô formula:

$$\begin{aligned}
C_{\lambda,\lambda'} &= \mathrm{Cov}_{a_0}\left[\left(\int_0^T A(t)\,dW(t)\right)^2, \left(\int_0^T B(t)\,dW(t)\right)^2\right] \\
&= \mathrm{Cov}_{a_0}\Bigg[2\int_0^T \left(\int_0^t A(s)\,dW(s)\right) A(t)\,dW(t) + \int_0^T A(t)^2\,dt, \\
&\qquad 2\int_0^T \left(\int_0^t B(s)\,dW(s)\right) B(t)\,dW(t) + \int_0^T B(t)^2\,dt\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&\leq 4 \operatorname{Cov}_{a_0}\left[\int_0^T \left(\int_0^t A(s)\, dW(s)\right) A(t)\, dW(t),\right.\\
&\qquad\qquad \left.\int_0^T \left(\int_0^t B(s)\, dW(s)\right) B(t)\, dW(t)\right]\\
&\quad + 2 \operatorname{Var}_{a_0}\left[\int_0^T \left(\int_0^t A(s)\, dW(s)\right) A(t)\, dW(t)\right]^{1/2} \operatorname{Var}_{a_0}\left[\int_0^T B(t)^2\, dt\right]^{1/2}\\
&\quad + 2 \operatorname{Var}_{a_0}\left[\int_0^T A(t)^2\, dt\right]^{1/2} \operatorname{Var}_{a_0}\left[\int_0^T \left(\int_0^t B(s)\, dW(s)\right) B(t)\, dW(t)\right]^{1/2}\\
&\quad + \operatorname{Var}_{a_0}\left[\int_0^T A(t)^2\, dt\right]^{1/2} \operatorname{Var}_{a_0}\left[\int_0^T B(t)^2\, dt\right]^{1/2}\\
&=: 4S_1(\lambda,\lambda') + 2S_2(\lambda,\lambda') + 2S_3(\lambda,\lambda') + S_4(\lambda,\lambda').
\end{aligned}
$$

We shall see that the main contribution comes from $S_1(\lambda,\lambda')$ and is of the form $T^2\langle Q_{a_0}\psi_\lambda,\psi_{\lambda'}\rangle^2$. We shall thus recover the terms appearing in the model problem.

The second estimate in Proposition 6 yields

$$
\operatorname{Var}_{a_0}\left[\int_0^T A(t)^2\, dt\right]^{1/2} = \mathbb{E}_{a_0}[\langle (Q_T - TQ_{a_0})\psi_\lambda,\psi_\lambda\rangle^2]^{1/2} \lesssim T^{1/2}2^{-2|\lambda|}
$$

and the corresponding estimate for $B(t)$ and λ'. The Itô isometry and the estimate (8.2.1) imply for the second kind of variance term the bound

$$
\begin{aligned}
&\operatorname{Var}_{a_0}\left[\int_0^T \left(\int_0^t A(s)\, dW(s)\right) A(t)\, dW(t)\right]^{1/2}\\
&= \left(\int_0^T \mathbb{E}_{a_0}\left[\left(\int_0^t A(s)\, dW(s)\right)^2 A(t)^2\right] dt\right)^{1/2}\\
&\leq \left(\int_0^T \mathbb{E}_{a_0}\left[\left(\int_0^t A(s)\, dW(s)\right)^4\right]^{1/2} \mathbb{E}_{a_0}[A(t)^4]^{1/2}\, dt\right)^{1/2}\\
&\lesssim \left(\int_0^T t\, \mathbb{E}_{a_0}[A(t)^4]\, dt\right)^{1/2}\\
&\lesssim T 2^{-2|\lambda|}
\end{aligned}
$$

and the corresponding result for the term involving $B(t)$. Let us use the estimates, we have so far obtained, to bound the sum of S_2, S_3 and S_4:

$$
\begin{aligned}
\operatorname{Var}_{a_0}[\mathcal{T}] &= \frac{1}{T^2}\sum_{|\lambda|,|\lambda'|\leq J(T)} 2^{4(|\lambda|+|\lambda'|)} C_{\lambda,\lambda'}\\
&\lesssim \sum_{|\lambda|,|\lambda'|\leq J(T)} T^{-2}2^{4(|\lambda|+|\lambda'|)}S_1(\lambda,\lambda') + (T^{-1}+T^{-1/2})2^{2(|\lambda|+|\lambda'|)}\\
&\sim T^{-1/2}2^{6J(T)} + \frac{1}{T^2}\sum_{|\lambda|,|\lambda'|\leq J(T)} 2^{4(|\lambda|+|\lambda'|)}S_1(\lambda,\lambda')\\
&\lesssim 2^{5J(T)} + \frac{1}{T^2}\sum_{|\lambda|,|\lambda'|\leq J(T)} 2^{4(|\lambda|+|\lambda'|)}S_1(\lambda,\lambda'). \qquad (8.2.2)
\end{aligned}
$$

It thus remains to estimate $S_1(\lambda,\lambda')$. We decompose $S_1(\lambda,\lambda')$, using the Itô isometry:

$$\begin{aligned}
S_1(\lambda,\lambda') &= \int_0^T \mathbb{E}_{a_0}\left[\left(\int_0^t A(s)\,dW(s)\right)A(t)\left(\int_0^t B(s)\,dW(s)\right)B(t)\right]dt\\
&= \int_0^T \left(\int_0^t \mathbb{E}_{a_0}[A(s)B(s)]\,ds\right)\mathbb{E}_{a_0}[A(t)B(t)]\,dt\\
&\quad + \int_0^T \operatorname{Cov}_{a_0}\left[\left(\int_0^t A(s)\,dW(s)\right)\left(\int_0^t B(s)\,dW(s)\right), A(t)B(t)\right]dt\\
&=: \int_0^T (I_1(t)+I_2(t))\,dt.
\end{aligned}$$

The first integrand $I_1(t)$ is immediately identified as $I_1(t) = t\langle Q_{a_0}\psi_\lambda, \psi_{\lambda'}\rangle^2$. Intuitively, the second integrand $I_2(t)$ is not growing linearly in t, since $A(t)$ and $B(t)$ are almost independent of $\mathcal{F}_s$ for times $s \ll t$ due to the mixing property of X.

For the exact argument we set $\tau := \max(0, \frac{2}{\delta}\log(t))$ and consider first the covariance with the integrals up to time $t-\tau$. Use Theorem 3 in the context of the subsequent remark about complete regularity (2.3.17) and apply again (8.2.1) to obtain

$$\begin{aligned}
&\operatorname{Cov}_{a_0}\left[\left(\int_0^{t-\tau} A(s)\,dW(s)\right)\left(\int_0^{t-\tau} B(s)\,dW(s)\right), A(t)B(t)\right]\\
&\lesssim e^{-\tau\delta/2}\operatorname{Var}_{a_0}\left[\left(\int_0^{t-\tau} A(s)\,dW(s)\right)\left(\int_0^{t-\tau} B(s)\,dW(s)\right)\right]^{1/2}\operatorname{Var}_{a_0}[A(t)B(t)]^{1/2}\\
&\lesssim t^{-1}(t-\tau)\,\mathbb{E}_{a_0}[A(0)^4]^{1/2}\,\mathbb{E}_{a_0}[B(0)^4]^{1/2}\\
&\lesssim 2^{-2(|\lambda|+|\lambda'|)}.
\end{aligned}$$

The remainder term for $I_2(t)$ can easily be estimated by the Cauchy-Schwarz inequality, since

$$\operatorname{Var}_{a_0}[A(t)B(t)]^{1/2} \le \mathbb{E}_{a_0}[A(t)^4]^{1/4}\,\mathbb{E}_{a_0}[B(t)^4]^{1/4} \lesssim 2^{-(|\lambda|+|\lambda'|)}, \quad t \ge 0,$$

and for $t \ge 1$

$$\begin{aligned}
&\operatorname{Var}_{a_0}\left[\left(\int_0^t A(s)\,dW(s)\right)\left(\int_0^t B(s)\,dW(s)\right)\right.\\
&\qquad\left. - \left(\int_0^{t-\tau} A(s)\,dW(s)\right)\left(\int_0^{t-\tau} B(s)\,dW(s)\right)\right]^{1/2}\\
&\le \operatorname{Var}_{a_0}\left[\left(\int_0^t A(s)\,dW(s)\right)\left(\int_{t-\tau}^t B(s)\,dW(s)\right)\right]^{1/2}\\
&\quad + \operatorname{Var}_{a_0}\left[\left(\int_{t-\tau}^t A(s)\,dW(s)\right)\left(\int_0^{t-\tau} B(s)\,dW(s)\right)\right]^{1/2}\\
&\le \mathbb{E}_{a_0}\left[\left(\int_0^t A(s)\,dW(s)\right)^4\right]^{1/4}\mathbb{E}_{a_0}\left[\left(\int_{t-\tau}^t B(s)\,dW(s)\right)^4\right]^{1/4}\\
&\quad + \mathbb{E}_{a_0}\left[\left(\int_{t-\tau}^t A(s)\,dW(s)\right)^4\right]^{1/4}\mathbb{E}_{a_0}\left[\left(\int_0^{t-\tau} B(s)\,dW(s)\right)^4\right]^{1/4}\\
&\lesssim (t\log(t))^{1/2}2^{-(|\lambda|+|\lambda'|)}
\end{aligned}$$

follow from the bound (8.2.1).

Resuming the calculations from the interim result (8.2.2), we finally obtain the announced bound (cf. the calculations for the model problem):

$$\begin{aligned}
\operatorname{Var}_{a_0}[\mathcal{T}] &\lesssim 2^{5J(T)} + \frac{1}{T^2} \sum_{|\lambda|,|\lambda'|\le J(T)} 2^{4(|\lambda|+|\lambda'|)} S_1(\lambda,\lambda') \\
&\lesssim 2^{5J(T)} + \sum_{|\lambda|,|\lambda'|\le J(T)} 2^{2(|\lambda|+|\lambda'|)} \frac{1}{T^2}\int_0^T (1+(t\log(t))^{1/2})\,dt \\
&\quad + \sum_{|\lambda|\le J(T)} 2^{4|\lambda|} \sum_{|\lambda'|\le J(T)} 2^{4|\lambda'|} \langle Q_{a_0}\psi_\lambda, \psi_{\lambda'}\rangle^2 \\
&\lesssim 2^{5J(T)} + T^{-1/2}\log(T)^{1/2} 2^{6J(T)} + \sum_{|\lambda|\le J(T)} 2^{4|\lambda|} \|Q_{a_0}\psi_\lambda\|^2_{H^2} \\
&\lesssim 2^{5J(T)} + \sum_{|\lambda|\le J(T)} 2^{4|\lambda|} \|\psi_\lambda\|^2_{L^2} \\
&\sim 2^{5J(T)}.
\end{aligned}$$

□

Remark 9. *Note that the constants appearing in the bounds of the preceding Proposition 14 can be chosen uniformly for all $a_0 \in M(R,\delta)$, $R > 0$ and $\delta > 0$ fixed, because the dependence on a_0 enters through the estimates for q_{a_0}, $v_0(a_0)$ and through the mixing property from Proposition 3, which all allow to choose uniform constants.*

Proposition 15. *For $a \in M^- \cap \mathcal{W}^s_{2,2}$, $s > 0$, the following estimates hold:*

$$\mathbb{E}_a[\mathcal{T} - \mathbb{E}_{a_0}[\mathcal{T}]] \ge C_1 T\|a-a_0\|^2_{\mathcal{W}^0_{2,2}} - C_2 T 2^{-2J(T)s}\|a-a_0\|^2_{\mathcal{W}^s_{2,2}} - C_3 2^{5J(T)/2},$$
$$\operatorname{Var}_a[\mathcal{T}] \le C_4(2^{5J(T)} + 2^{2J(T)}T\|a-a_0\|^2_{\mathcal{W}^0_{2,2}})(1+\|a-a_0\|^2_{\mathcal{W}^0_{2,2}}).$$

The positive constants C_1, C_2, C_3, C_4 can be chosen uniformly for all $a \in M(R,\delta)$, when $R > 0$ and $\delta > 0$ are fixed.

Proof. In the course of the proof of the bound for the expected value we shall number the appearing constants consecutively, because the use of the symbol $\lesssim$ is not applicable for differences. Note that eventually the constants C_3, C_4 and C_7+C_9 correspond to the constants C_1, C_2 and C_3 in the statement of the proposition. Starting with the expected value, we first look at our model problem and then transfer the results to the more complicated SDDE case.

▷ **Model problem.** The mapping properties of Q on $\mathcal{W}^s_{2,2}$ (Theorem 2) and the Jackson inequality in H^2 (Remark 11) yield

$$\begin{aligned}
\mathbb{E}_a[\mathcal{T} - \mathbb{E}_{a_0}[\mathcal{T}]] &= \sum_{|\lambda|\le J(T)} 2^{4|\lambda|}(\mathbb{E}[\langle T^{1/2}Q(a-a_0)+\Gamma, \psi_\lambda\rangle^2] - \langle Q\psi_\lambda, \psi_\lambda\rangle^2) \\
&= \sum_{|\lambda|\le J(T)} 2^{4|\lambda|} T\langle Q(a-a_0), \psi_\lambda\rangle^2 \\
&= T(C_1\|Q(a-a_0)\|^2_{H^2} - C_2\|(\mathrm{Id} - P_{J(T)})Q(a-a_0)\|^2_{H^2}) \\
&\ge C_3 T\|a-a_0\|^2_{\mathcal{W}^0_{2,2}} - C_4 T 2^{-2J(T)s}\|Q(a-a_0)\|^2_{H^{s+2}} \\
&\ge C_3 T\|a-a_0\|^2_{\mathcal{W}^0_{2,2}} - C_5 T 2^{-2J(T)s}\|a-a_0\|^2_{\mathcal{W}^s_{2,2}}.
\end{aligned}$$

The constants C_1, C_2 and C_4 only depend on the wavelet basis, whereas C_3 and C_5 depend also on the operator norm of $Q : \mathcal{W}^s_{2,2} \to H^{s+2}([-r,0])$ and its inverse.

In the SDDE case certain terms do not cancel:

$$\begin{aligned}
&\mathbb{E}_a[\mathcal{T} - \mathbb{E}_{a_0}[\mathcal{T}]] \\
&= \frac{1}{T} \sum_{|\lambda| \le J(T)} 2^{4|\lambda|} \Big(\mathbb{E}_a[(\langle b_T - Q_T a, \psi_\lambda\rangle + \langle Q_T(a-a_0), \psi_\lambda\rangle)^2] - \langle T Q_{a_0} \psi_\lambda, \psi_\lambda\rangle \Big) \\
&= \frac{1}{T} \sum_{|\lambda| \le J(T)} 2^{4|\lambda|} \Big(\langle T(Q_a - Q_{a_0})\psi_\lambda, \psi_\lambda\rangle + 2\,\mathbb{E}_a[\langle b_T - Q_T a, \psi_\lambda\rangle \langle Q_T(a-a_0), \psi_\lambda\rangle] \\
&\qquad\qquad + \mathbb{E}_a[\langle Q_T(a-a_0), \psi_\lambda\rangle^2] \Big) \\
&=: \frac{1}{T} \sum_{|\lambda| \le J(T)} 2^{4|\lambda|} (S_1(\lambda) + 2S_2(\lambda) + S_3(\lambda)).
\end{aligned}$$

It will become apparent that the last summand $S_3(\lambda)$ gives the main contribution. For the sum over $S_3(\lambda)$, the inequality $\mathbb{E}[X^2] \ge \mathbb{E}[X]^2$ yields the same expression as lower bound as in the model problem:

$$\begin{aligned}
\sum_{|\lambda| \le J(T)} 2^{4|\lambda|} S_3(\lambda) &\ge \sum_{|\lambda| \le J(T)} 2^{4|\lambda|} \langle T Q_a(a-a_0), \psi_\lambda\rangle^2 \\
&\ge C_3 \|a - a_0\|^2_{\mathcal{W}^0_{2,2}} - C_5 2^{-2J(T)s} \|a - a_0\|^2_{\mathcal{W}^s_{2,2}}.
\end{aligned}$$

Recall that C_3 and C_5 depend on the operator norms of Q_a and Q_a^{-1}. By Corollary 7 they can be chosen uniformly for $a \in M(R, \delta)$.

From Corollary 13 and $q_a - q_{a_0} \in H^2([-r, r])$ (Proposition 1) we conclude for the sum over $S_1(\lambda)$

$$\sum_{|\lambda| \le J(T)} 2^{4|\lambda|} S_1(\lambda) \ge -C_6 \sum_{|\lambda| \le J(T)} 2^{4|\lambda|} 2^{-|\lambda|(2+1-\frac{1}{2})} T \|q_a - q_{a_0}\|_{H^2} \ge -C_7 2^{5J(T)/2} T.$$

The constants C_6 and C_7 depend $\mathcal{W}^0_{2,2}$-continuously on a by Proposition 4.

For estimating $S_2(\lambda)$ we use $\mathbb{E}_a[\langle b_T - Q_T a, \psi_\lambda\rangle] = 0$, the Cauchy-Schwarz inequality, Proposition 6 and Corollaries 3 and 13:

$$\begin{aligned}
\sum_{|\lambda| \le J(T)} 2^{4|\lambda|} S_2(\lambda) &= \sum_{|\lambda| \le J(T)} 2^{4|\lambda|}\, \mathbb{E}_a[\langle (Q_T - T Q_a)(a - a_0), \psi_\lambda\rangle \langle b_T - Q_T a, \psi_\lambda\rangle] \\
&\ge -C_8 \sum_{|\lambda| \le J(T)} 2^{4|\lambda|} \|a - a_0\|_{TV} T 2^{-3|\lambda|/2} \langle Q_a \psi_\lambda, \psi_\lambda\rangle^{1/2} \\
&\ge -C_9 2^{5J(T)/2} T.
\end{aligned}$$

By the Proposition and Corollaries used, the constants can be chosen uniformly for $a \in M(R, \delta)$. By combining the single estimates we find

$$\mathbb{E}_a[\mathcal{T} - \mathbb{E}_{a_0}[\mathcal{T}]] \ge C_3 T \|a - a_0\|^2_{\mathcal{W}^0_{2,2}} - C_5 T 2^{-2J(T)s} \|a - a_0\|^2_{\mathcal{W}^s_{2,2}} - (C_7 + C_9) 2^{5J(T)/2}$$

with uniform constants for $a \in M(R, \delta)$.

The estimate of the variance of $\mathcal{T}$ under $\mathbb{P}_a$ is more involved. So we start off again with the model problem.

▷ **Model problem.** We expand the covariance into four terms and use the fact that odd moments of Gaussian vectors vanish:

$$\begin{aligned}
\mathrm{Var}_a[\mathcal{T}] = \sum_{|\lambda|,|\lambda'|\le J(T)} & 2^{4(|\lambda|+|\lambda'|)}\,\mathrm{Cov}\Big[\langle T^{1/2}Q(a-a_0)+\Gamma,\psi_\lambda\rangle^2, \\
& \qquad\qquad \langle T^{1/2}Q(a-a_0)+\Gamma,\psi_{\lambda'}\rangle^2\Big] \\
= \mathrm{Var}_{a_0}[\mathcal{T}] + & \sum_{|\lambda|,|\lambda'|\le J(T)} 2^{4(|\lambda|+|\lambda'|)}\Big(\\
& 4T\langle Q(a-a_0),\psi_\lambda\rangle\langle Q(a-a_0),\psi_{\lambda'}\rangle\,\mathrm{Cov}[\langle\Gamma,\psi_\lambda\rangle,\ \langle\Gamma,\psi_{\lambda'}\rangle] \\
& + 2T^{1/2}\langle Q(a-a_0),\psi_\lambda\rangle\,\mathrm{Cov}[\langle\Gamma,\psi_\lambda\rangle,\ \langle\Gamma,\psi_{\lambda'}\rangle^2] \\
& + 2T^{1/2}\langle Q(a-a_0),\psi_{\lambda'}\rangle\,\mathrm{Cov}[\langle\Gamma,\psi_\lambda\rangle^2,\ \langle\Gamma,\psi_{\lambda'}\rangle]\Big) \\
= \mathrm{Var}_{a_0}[\mathcal{T}] + & \\
\sum_{|\lambda|,|\lambda'|\le J(T)} & 2^{4(|\lambda|+|\lambda'|)}4T\langle Q(a-a_0),\psi_\lambda\rangle\langle Q(a-a_0),\psi_{\lambda'}\rangle\langle Q\psi_\lambda,\psi_{\lambda'}\rangle \\
\le \mathrm{Var}_{a_0}[\mathcal{T}] + 4T & \Big\|\sum_{|\lambda|\le J(T)} 2^{4|\lambda|}\langle Q(a-a_0),\psi_\lambda\rangle Q\psi_\lambda\Big\|_{L^2}\bullet \\
& \bullet\Big\|\sum_{|\lambda|\le J(T)} 2^{4|\lambda|}\langle Q(a-a_0),\psi_\lambda\rangle\psi_\lambda\Big\|_{L^2}.
\end{aligned}$$

Theorem 5 yields for $f \in L^2([-r,0])$ the duality result

$$\|Qf\|_{L^2} = \sup_{\|g\|_{L^2}=1}\langle Qf,g\rangle = \sup_{\|g\|_{L^2}=1}\langle f,Qg\rangle \lesssim \sup_{\|h\|_{H^2}=1}\langle f,h\rangle = \|f\|_{H^{-2}}.$$

The characterisation of H^{-2} and H^2 by the wavelet basis (Appendix A.3) thus implies

$$\begin{aligned}
\mathrm{Var}_a[\mathcal{T}] \lesssim \mathrm{Var}_{a_0}[\mathcal{T}] + T2^{2J(T)} & \Big\|\sum_{|\lambda|\le J(T)} 2^{4|\lambda|}\langle Q(a-a_0),\psi_\lambda\rangle\psi_\lambda\Big\|_{H^{-2}}\bullet \\
& \bullet\Big\|\sum_{|\lambda|\le J(T)} 2^{2|\lambda|}\langle Q(a-a_0),\psi_\lambda\rangle\psi_\lambda\Big\|_{L^2} \\
\lesssim \mathrm{Var}_{a_0}[\mathcal{T}] + 2^{2J(T)}T & \sum_{|\lambda|\le J(T)} 2^{4|\lambda|}\langle Q(a-a_0),\psi_\lambda\rangle^2 \\
\lesssim \mathrm{Var}_{a_0}[\mathcal{T}] + 2^{2J(T)}T & \|Q(a-a_0)\|_{H^2}^2 \\
\sim \mathrm{Var}_{a_0}[\mathcal{T}] + 2^{2J(T)}T & \|a-a_0\|_{\mathcal{W}^0_{2,2}}^2.
\end{aligned}$$

The constant involved depends continuously on the operator norm of $Q : \mathcal{W}^0_{2,2} \to H^2$ and its inverse.

In the SDDE case we decompose the variance of $\mathcal{T}$ using the general inequality

$\mathrm{Var}[X_1+\cdots+X_n] \le n(\mathrm{Var}[X_1]+\cdots+\mathrm{Var}[X_n])$:

$$\begin{aligned}\mathrm{Var}_a[\mathcal{T}] &= \mathrm{Var}_a\left[\frac{1}{T}\sum_{|\lambda|\le J(T)} 2^{4|\lambda|}(\langle b_T-Q_Ta,\psi_\lambda\rangle+\langle Q_T(a-a_0),\psi_\lambda\rangle)^2\right]\\
&\le 3\,\mathrm{Var}_a\left[\frac{1}{T}\sum_{|\lambda|\le J(T)} 2^{4|\lambda|}\langle b_T-Q_Ta,\psi_\lambda\rangle^2\right]\\
&\quad + 12\,\mathrm{Var}_a\left[\frac{1}{T}\sum_{|\lambda|\le J(T)} 2^{4|\lambda|}\langle Q_T(a-a_0),\psi_\lambda\rangle\langle b_T-Q_Ta,\psi_\lambda\rangle\right] \qquad (8.2.3)\\
&\quad + 3\,\mathrm{Var}_a\left[\frac{1}{T}\sum_{|\lambda|\le J(T)} 2^{4|\lambda|}\langle Q_T(a-a_0),\psi_\lambda\rangle^2\right].\end{aligned}$$

The first variance term is the analogue of $\mathrm{Var}_{a_0}[\mathcal{T}]$ estimated in Proposition 14, only with a_0 replaced by a. Hence taking into account Remark 9, we find the uniform bound for $a\in M(R,\delta)$ of the first variance term:

$$\mathrm{Var}_a\left[\frac{1}{T}\sum_{|\lambda|\le J(T)} 2^{4|\lambda|}\langle b_T-Q_Ta,\psi_\lambda\rangle^2\right] \lesssim 2^{5J(T)}.$$

The second variance term in (8.2.3) resembles the mixed term in the model problem apart from the randomness of Q_T. Therefore this term is decomposed further. The first term involves the difference $Q_T - TQ_a$ which can be estimated using Proposition 6 and Lemma 8 with $x^4 \lesssim e^x$, $x\ge 0$. The second term corresponds exactly to the mixed term in the model problem, since only second order moments of $b_T - Q_Ta$ are involved, so that we just use the bound derived there and obtain

$$\begin{aligned}&\mathrm{Var}_a\left[\frac{1}{T}\sum_{|\lambda|\le J(T)} 2^{4|\lambda|}\langle Q_T(a-a_0),\psi_\lambda\rangle\langle b_T-Q_Ta,\psi_\lambda\rangle\right]\\
&\le 2\,\mathrm{Var}_a\left[\sum_{|\lambda|\le J(T)} 2^{4|\lambda|}\langle(\tfrac{1}{T}Q_T-Q_a)(a-a_0),\psi_\lambda\rangle\langle b_T-Q_Ta,\psi_\lambda\rangle\right]\\
&\qquad + 2\,\mathrm{Var}_a\left[\sum_{|\lambda|\le J(T)} 2^{4|\lambda|}\langle Q_a(a-a_0),\psi_\lambda\rangle\langle b_T-Q_Ta,\psi_\lambda\rangle)\right]\\
&\le 2\,2^{J(T)}\sum_{|\lambda|\le J(T)} 2^{8|\lambda|}\,\mathbb{E}_a[\langle(\tfrac{1}{T}Q_T-Q_a)(a-a_0),\psi_\lambda\rangle^4]^{1/2}\,\mathbb{E}_a[\langle b_T-Q_Ta,\psi_\lambda\rangle^4]^{1/2}\\
&\quad + 2\,\mathrm{Var}_a\left[\sum_{|\lambda|\le J(T)} 2^{4|\lambda|}\langle Q_a(a-a_0),\psi_\lambda\rangle\langle b_T-Q_Ta,\psi_\lambda\rangle\right]\\
&\lesssim 2^{J(T)}\sum_{|\lambda|\le J(T)}\left(2^{8|\lambda|}T^{-1}\|a-a_0\|_{TV}^2 2^{-3|\lambda|}T2^{-2|\lambda|}\right) + 2^{2J(T)}T\|a-a_0\|_{\mathcal{W}^0_{2,2}}^2\\
&\lesssim (2^{5J(T)}+2^{2J(T)}T)\|a-a_0\|_{\mathcal{W}^0_{2,2}}^2.\end{aligned}$$

The last variance term in (8.2.3) is treated by writing $Q_T = (Q_T - TQ_a) + TQ_a$ and by separately regarding the second moments of the two generated summands. We deduce from Proposition 6, the Cauchy-Schwarz inequality and the 2-regularity of the wavelet basis

$$\begin{aligned}
&\operatorname{Var}_a\left[\frac{1}{T}\sum_{|\lambda|\le J(T)} 2^{4|\lambda|}\langle Q_T(a-a_0),\psi_\lambda\rangle^2\right]\\
&=\operatorname{Var}_a\Bigg[\frac{1}{T}\sum_{|\lambda|\le J(T)} 2^{4|\lambda|}\Big(\langle (Q_T-TQ_a)(a-a_0),\psi_\lambda\rangle^2\\
&\qquad\qquad + 2\langle TQ_a(a-a_0),\psi_\lambda\rangle\langle Q_T(a-a_0),\psi_\lambda\rangle\Big)\Bigg]\\
&\le 2\,\mathbb{E}_a\left[\left(\frac{1}{T}\sum_{|\lambda|\le J(T)} 2^{4|\lambda|}\langle (Q_T-TQ_a)(a-a_0),\psi_\lambda\rangle^2\right)^2\right]\\
&\quad + 8\,\mathbb{E}_a\left[\left(\frac{1}{T}\sum_{|\lambda|\le J(T)} 2^{4|\lambda|}\langle TQ_a(a-a_0),\psi_\lambda\rangle\langle (Q_T-TQ_a)(a-a_0),\psi_\lambda\rangle\right)^2\right]\\
&\lesssim \frac{2^{J(T)}}{T^2}\sum_{|\lambda|\le J(T)} 2^{8|\lambda|}\,\mathbb{E}_a[\langle (Q_T-TQ_a)(a-a_0),\psi_\lambda\rangle^4]\\
&\quad + \frac{2^{J(T)}}{T^2}\sum_{|\lambda|\le J(T)} 2^{8|\lambda|}\langle TQ_a(a-a_0),\psi_\lambda\rangle^2\,\mathbb{E}_a[\langle (Q_T-TQ_a)(a-a_0),\psi_\lambda\rangle^2]\\
&\lesssim \frac{2^{J(T)}}{T^2}\sum_{|\lambda|\le J(T)} 2^{8|\lambda|}(T^2 2^{-6|\lambda|} + T^2\langle Q_a(a-a_0),\psi_\lambda\rangle^2 T 2^{-3|\lambda|}\|a-a_0\|_{TV}^2)\\
&\lesssim 2^{4J(T)} + T2^{2J(T)}\|Q_a(a-a_0)\|_{H^2}^2\|a-a_0\|_{\mathcal{W}_{2,2}^0}^2\\
&\lesssim 2^{4J(T)} + 2^{2J(T)}T\|a-a_0\|_{\mathcal{W}_{2,2}^0}^4.
\end{aligned}$$

By Theorem 5 and Proposition 6 the constants can again be chosen uniformly for $a\in M(R,\delta)$.

We have thus proved the announced asymptotic upper bound

$$\operatorname{Var}_a[\mathcal{T}] \lesssim (2^{5J(T)} + 2^{2J(T)}T\|a-a_0\|_{\mathcal{W}_{2,2}^0}^2)(1+\|a-a_0\|_{\mathcal{W}_{2,2}^0}^2)$$

with a uniform constant for $a\in M(R,\delta)$. □

We are now ready to prove the main theorem of this section.

Theorem 9. *Given the decision function φ from Definition 13 and some fixed $\beta\in(0,1)$, we can choose $\eta>0$ such that for $\rho(T):=\eta T^{-\frac{s}{2s+2.5}}$ holds*

$$\limsup_{T\to\infty}\sup_{a\in\mathcal{K}_{\rho(T)}}\mathbb{P}_a(\varphi=0)\le\beta.$$

Proof. Set $b:=\mathbb{E}_a[\mathcal{T}]-\mathbb{E}_{a_0}[\mathcal{T}]-q_{1-\alpha}\operatorname{Var}_{a_0}[\mathcal{T}]^{1/2}$. Then we infer from the preceding Propositions 14 and 15

$$b\ge C_1T\|a-a_0\|_{\mathcal{W}_{2,2}^0}^2 - C_2T2^{-2J(T)s}S^2 - C_3 2^{5J(T)/2} - q_{1-\alpha}C_4 2^{5J(T)/2}$$

uniformly for all $a\in\mathcal{K}_\rho\subset M(S+\|a_0\|,\delta)$ with positive constants C_1,C_2,C_3,C_4. Due to $2^{-2J(T)s}\sim T^{-1}2^{5J(T)/2}\sim T^{-\frac{2s}{2s+2.5}}$ we can choose the constant η so large that for all $a \in \mathcal{K}_\rho$ and sufficiently large T the inequality $b \geq C_5 T \|a - a_0\|^2_{\mathcal{W}^0_{2,2}}$ is valid with $C_5 = C_5(C_1, C_2, C_3, C_4, \eta) > 0$. We deduce

$$
\begin{aligned}
&\mathbb{P}_a(\varphi = 0)\\
&= \mathbb{P}_a\left(\frac{\mathcal{T} - \mathbb{E}_{a_0}[\mathcal{T}]}{\operatorname{Var}_{a_0}[\mathcal{T}]^{1/2}} \leq q_{1-\alpha}\right)\\
&= \mathbb{P}_a(-\mathcal{T} + \mathbb{E}_a[\mathcal{T}] \geq b)\\
&\leq \frac{\operatorname{Var}_a[\mathcal{T}]}{b^2}\\
&\leq C_1 \frac{(2^{5J(T)} + 2^{2J(T)} T \|a - a_0\|^2_{\mathcal{W}^0_{2,2}})(1 + \|a - a_0\|^2_{\mathcal{W}^0_{2,2}})}{C_5^2 T^2 \|a - a_0\|^4_{\mathcal{W}^0_{2,2}}}\\
&\leq C_6 T^{-\frac{4s}{2s+2.5}} \rho(T)^{-4} + C_7 T^{-\frac{2s+0.5}{2s+2.5}} \rho(T)^{-2} + C_8 T^{-\frac{2s}{2s+2.5}} \rho(T)^{-2} + C_9 T^{-\frac{2s+0.5}{2s+2.5}}\\
&\leq C_6 \eta^{-4} + (C_7 + C_8)\eta^{-2} + C_9 T^{-\frac{1}{5}}
\end{aligned}
$$

with positive constants C_6, C_7, C_8, C_9. Enlarging η if necessary, this last sum can obviously be bounded by any $\beta > 0$ for sufficiently large values of T. Since the constants do not depend on $a \in \mathcal{K}_\rho$, this bound holds uniformly in a. $\square$

Remark 10. *Looking at the results of the preceding chapter, it would be interesting to develop an adaptive testing method for nonparametric alternatives of some Besov regularity. In the classical signal plus white noise model such a test has been developed for instance by Spokoiny (1996). It seems reasonable that the adaptive techniques yield comparable results in the case of the SDDE problem. However, the technical issues regarding the estimates will become even more unpleasant. A similar statement applies to the problem of testing not a single, but a parametric hypothesis against a nonparametric alternative, which in a regression setting has been studied by Horowitz and Spokoiny (1999) among many others.*

8.3 Optimality of the test

We shall establish the counterpart to the theory of the preceding section, namely an asymptotic lower bound of the separation rate for any testing procedure for $T \to \infty$. The technique employed is similar to that of the proof of the lower bound for the estimation problem. Its analogue for the signal plus white noise model is presented by Ingster (1993b).

Recall that a non-randomized decision function of a test based on the observations of X on $[0, T]$ is just an $\mathcal{F}^X_T$-measurable function with values in the set $\{0, 1\}$. Therefore the next theorem states that the separation rate $\rho(T)$ of hypothesis and alternative cannot be asymptotically smaller than $T^{-\frac{s}{2s+2.5}}$ for the uniform testing problem, whichever decision function is used.

Theorem 10. *For the testing problem specified in Definition 12, but with $s > \frac{1}{4}$, there exists for any given $\eta > 0$ a constant $C_\eta > 0$ such that for $\rho(T) := C_\eta T^{-\frac{s}{2s+2.5}}$ the asymptotic lower bound*

$$
\liminf_{T\to\infty} \inf_{\Phi} \sup_{a \in \mathcal{K}_{\rho(T)}} \big(\mathbb{P}_{a_0}(\Phi = 1) + \mathbb{P}_a(\Phi = 0)\big) \geq 1 - \eta
$$

holds, where the infimum is taken over all $\{0, 1\}$-valued and $\mathcal{F}^X_T$-measurable functions Φ.

Proof. Let (ψ_{jk}) be a compactly supported $(s \vee 2)$-regular wavelet basis of $L^2(\mathbb{R})$ and denote by R_j a maximal subset of $\mathbb{Z}$ with $\operatorname{supp}(\psi_{jk}) \subset [-r, 0]$ and $\operatorname{supp}(\psi_{jk}) \cap \operatorname{supp}(\psi_{jk'}) = \varnothing$ for all $k, k' \in R_j$, $k \neq k'$. For $\varepsilon \in \{-1, +1\}^{R_j}$ and $\gamma > 0$ put

$$d_\varepsilon := \gamma \sum_{k \in R_j} \varepsilon_k \psi_{jk}.$$

Choose $\gamma := \gamma_0 2^{-j(s+\frac{1}{2})}$ with some sufficiently small constant $\gamma_0 > 0$ and $2^j \sim T^{\frac{1}{2s+2.5}}$ such that $\|d_\varepsilon\|_s \leq S$ holds, which is possible due to $\|d_\varepsilon\|_s \sim \gamma 2^{j(s+\frac{1}{2})}$. In the sequel, the constants that appear will be independent of γ_0. Putting $C_\eta := C_1\gamma_0$ with $C_1 > 0$ small enough we can guarantee that $a_\varepsilon := a_0 + d_\varepsilon$ eventually lies in $\mathcal{K}_{\rho(T)}$ for large T, because for all ε

$$\rho(T) = C_\eta T^{-\frac{s}{2s+2.5}} = C_1\gamma_0 T^{-\frac{s}{2s+2.5}} \leq C_1 C_2 \gamma 2^{j/2} \leq C_1 C_3 \liminf_{T\to\infty} \|d_\varepsilon\|_{L^2} \tag{8.3.4}$$

holds. Note that $v_0(a_0 + d_\varepsilon) \leq -\delta$ follows from Theorem 1 for large T, since $\|d_\varepsilon\|_{L^2} \to 0$ for $T \to \infty$ and $v_0(a_0) < -\delta$ are satisfied.

We bound the following sum R of probabilities from below (Φ denotes any decision function) by the usual Bayes technique of estimating the supremum over a set from below by the mean over this set:

$$\begin{aligned}
R &:= \sup_{\varepsilon \in \{-1,+1\}^{R_j}} \big(\mathbb{P}_{a_0}(\Phi = 1) + \mathbb{P}_{a_\varepsilon}(\Phi = 0)\big) \\
&= \sup_{\varepsilon \in \{-1,+1\}^{R_j}} \mathbb{E}_{a_0}\Big[\Phi + \tfrac{d\,\mathbb{P}_{a_\varepsilon}}{d\,\mathbb{P}_{a_0}}(1 - \Phi)\Big] \\
&\geq 2^{-|R_j|} \sum_{\varepsilon \in \{-1,+1\}^{R_j}} \mathbb{E}_{a_0}\Big[\Phi + \tfrac{d\,\mathbb{P}_{a_\varepsilon}}{d\,\mathbb{P}_{a_0}}(1 - \Phi)\Big] \\
&= 1 + \mathbb{E}_{a_0}\left[\Phi\left(1 - 2^{-|R_j|} \sum_{\varepsilon \in \{-1,+1\}^{R_j}} \frac{d\,\mathbb{P}_{a_\varepsilon}}{d\,\mathbb{P}_{a_0}}\right)\right] \\
&\geq 1 - \mathbb{E}_{a_0}[\Phi^2]^{1/2}\, \mathbb{E}_{a_0}\left[\left(1 - 2^{-|R_j|} \sum_{\varepsilon \in \{-1,+1\}^{R_j}} \frac{d\,\mathbb{P}_{a_\varepsilon}}{d\,\mathbb{P}_{a_0}}\right)^2\right]^{1/2} \\
&\geq 1 - \left(\mathbb{E}_{a_0}\left[\left(2^{-|R_j|} \sum_{\varepsilon \in \{-1,+1\}^{R_j}} \frac{d\,\mathbb{P}_{a_\varepsilon}}{d\,\mathbb{P}_{a_0}}\right)^2\right] - 1\right)^{1/2}.
\end{aligned} \tag{8.3.5}$$

We shall prove that the expected value in the last line is less than $\exp(C_6\gamma_0^4)$ with some uniform constant $C_6 > 0$. For a given $\eta > 0$ we may choose γ_0 with $(\exp(C_6\gamma_0^4) - 1)^{1/2} \leq \eta$ so that the error probability R satisfies $R \geq 1 - \eta$. After this choice of γ_0 we set $C_\eta := C_1\gamma_0$, where some $C_1 < C_3^{-1}$ is chosen (cf. (8.3.4)), and the theorem follows.

▷ **Model problem.** The likelihood ratio between the Gaussian distributions of the observations $Y_{a_\varepsilon} = Qa_\varepsilon + T^{-1/2}\Gamma$ and $Y_{a_0} = Qa_0 + T^{-1/2}\Gamma$ is given by Da Prato and Zabczyk (1992, Thm. 2.21)

$$\begin{aligned}
\frac{d\,\mathbb{P}_{a_\varepsilon}}{d\,\mathbb{P}_{a_0}}(Y) &= \exp(\langle T^{1/2}Q^{-1/2}Q(a_\varepsilon - a_0), T^{1/2}Q^{-1/2}Y\rangle)\bullet \\
&\quad \bullet \exp(-\tfrac{1}{2}\|T^{1/2}Q^{-1/2}Q(a_\varepsilon - a_0)\|^2) \\
&= \exp(T\langle d_\varepsilon, Y\rangle - \tfrac{T}{2}\|Q^{1/2}d_\varepsilon\|^2).
\end{aligned}$$

Let us here assume that Q is the covariance operator Q_W from Remark 5. Then it is claimed that for $k,k' \in R_j$ with $k \neq k'$ the orthogonality relation $\langle Q_W \psi_{jk}, \psi_{jk'}\rangle = 0$ is valid. In fact, using the property $\int \psi_{jk} = \int \psi_{jk'} = 0$ and the disjointness of the support intervals we obtain by partial integration

$$\begin{aligned}\langle Q_W \psi_{jk}, \psi_{jk'}\rangle &= \int_{-r}^{0}\int_{-r}^{0} \min(t,s)\psi_{jk}(t)\psi_{jk'}(s)\,dt\,ds \\ &= \int_{-r}^{0}\int_{t}^{0}\psi_{jk}(u)\,du \int_{t}^{0}\psi_{jk'}(v)\,dv\,dt = 0.\end{aligned}$$

We have used a well-known result on the reproducing kernel Hilbert space of Brownian motion.

Under the assumption $Q = Q_W$ the calculations are straightforward (the arguments are given afterwards and the constants numbered consecutively):

$$\begin{aligned}&\mathbb{E}_{a_0}\left[\left(2^{-|R_j|}\sum_{\varepsilon\in\{-1,+1\}^{R_j}}\frac{d\,\mathbb{P}_{a_\varepsilon}}{d\,\mathbb{P}_{a_0}}\right)^2\right] \\ &= 2^{-2|R_j|}\sum_{\varepsilon,\varepsilon'\in\{-1,+1\}^{R_j}}\mathbb{E}_{a_0}\Big[\exp\Big(T\langle d_\varepsilon + d_{\varepsilon'}, Y\rangle - \tfrac{T}{2}(\|Q_W^{1/2}d_\varepsilon\|^2 + \|Q_W^{1/2}d_{\varepsilon'}\|^2)\Big)\Big] \\ &\overset{(1)}{=} 2^{-2|R_j|}\sum_{\varepsilon,\varepsilon'\in\{-1,+1\}^{R_j}}\mathbb{E}_{a_0+d_\varepsilon+d'_\varepsilon}\Big[\exp\Big(T\langle Q_W^{1/2}d_\varepsilon, Q_W^{1/2}d_{\varepsilon'}\rangle\Big)\Big] \\ &= 2^{-2|R_j|}\sum_{\varepsilon,\varepsilon'\in\{-1,+1\}^{R_j}}\exp(T\langle Q_W d_\varepsilon, d_{\varepsilon'}\rangle) \\ &\overset{(2)}{=} 2^{-2|R_j|}\sum_{\varepsilon,\varepsilon'\in\{-1,+1\}^{R_j}}\exp\left(T\gamma^2\sum_{k\in R_j}\varepsilon_k\varepsilon'_k\langle Q_W\psi_{jk},\psi_{jk}\rangle\right) \\ &\overset{(3)}{=} 2^{-|R_j|}\sum_{\varepsilon\in\{-1,+1\}^{R_j}}\exp\left(T\gamma^2\sum_{k\in R_j}\varepsilon_k\langle Q_W\psi_{jk},\psi_{jk}\rangle\right) \\ &\overset{(4)}{=} \prod_{k\in R_j}\tfrac{1}{2}\Big(\exp(T\gamma^2\langle Q_W\psi_{jk},\psi_{jk}\rangle) + \exp(-T\gamma^2\langle Q_W\psi_{jk},\psi_{jk}\rangle)\Big) \\ &\overset{(5)}{\le} \cosh(C_4 T\gamma^2 2^{-2j})^{C_5 2^j} \\ &\overset{(6)}{\le} \exp(C_4^2 T^2\gamma^4 2^{-4j})^{C_5 2^j} \\ &\le \exp(C_6\gamma_0^4).\end{aligned}$$

Identity (1) follows from reinjecting the likelihood for $\mathbb{P}_{a_0+d_\varepsilon+d_{\varepsilon'}}$, (2) from the orthogonality relation just derived, (3) from the invariance of the sum over ε' for fixed ε, (4) from expanding the product, (5) from Corollaries 3 and 13 and $|R_j| \le C_5 2^j$, and (6) from the inequality $\cosh(x) \le e^{x^2}$, $x \in \mathbb{R}$.

So, for the model problem with $Q = Q_W$ we obtain the asserted bound for the expected value and the asymptotic lower bound follows.

From Corollary 2 and Remark 2, which we may apply like in the proof of Theorem 5, we infer for the SDDE case that the summands resulting from expanding the square of the sum equal, for $\varepsilon, \varepsilon' \in \{-1,+1\}^{R_j}$ and $T \geq r$,

$$
\begin{aligned}
&\mathbb{E}_{a_0}\left[\tfrac{d\mathbb{P}_{a_\varepsilon}}{d\mathbb{P}_{a_0}}\tfrac{d\mathbb{P}_{a_{\varepsilon'}}}{d\mathbb{P}_{a_0}}\right] \\
&= \mathbb{E}_{a_0}\Bigg[\Lambda_r(X^{(a_\varepsilon)}, X^{(a_0)})\Lambda_r(X^{(a_{\varepsilon'})}, X^{(a_0)}) \bullet \\
&\qquad \bullet \exp\left(\int_r^T \int_{-r}^0 X(t+s)\, d(a_\varepsilon + a_{\varepsilon'} - 2a_0)(s)\, dW(t)\right) \bullet \\
&\qquad \bullet \exp\left(-\frac{1}{2}\int_r^T \left(\int_{-r}^0 X(t+s)\, d(a_\varepsilon - a_0)(s)\right)^2 dt\right) \bullet \\
&\qquad \bullet \exp\left(-\frac{1}{2}\int_r^T \left(\int_{-r}^0 X(t+s)\, d(a_{\varepsilon'} - a_0)(s)\right)^2 dt\right)\Bigg] \\
&= \mathbb{E}_{a_0+d_\varepsilon+d_{\varepsilon'}}\Bigg[\exp\left(\int_r^T \left(\int_{-r}^0 X(t+s) d_\varepsilon(s)\, ds\right)\left(\int_{-r}^0 X(t+s) d_{\varepsilon'}(s)\, ds\right) dt\right) \bullet \\
&\quad \bullet \Lambda_r(X^{(a_\varepsilon)}, X^{(a_0)})\Lambda_r(X^{(a_{\varepsilon'})}, X^{(a_0)})\Lambda_r(X^{(a_0)}, X^{(a_0+d_\varepsilon+d_{\varepsilon'})})\Bigg] \\
&\leq \mathbb{E}_{a_0+d_\varepsilon+d_{\varepsilon'}}\left[\exp\left(2\int_r^T \left(\int_{-r}^0 X(t+s) d_\varepsilon(s)\, ds\right)\left(\int_{-r}^0 X(t+s) d_{\varepsilon'}(s)\, ds\right) dt\right)\right]^{1/2} \\
&\quad \bullet \mathbb{E}_{a_0+d_\varepsilon+d_{\varepsilon'}}\left[\Lambda_r(X^{(a_\varepsilon)}, X^{(a_0)})^2 \Lambda_r(X^{(a_{\varepsilon'})}, X^{(a_0+d_\varepsilon+d_{\varepsilon'})})^2\right]^{1/2}.
\end{aligned}
$$

Since $\|d_\varepsilon\|_{L^1} = \|d_{\varepsilon'}\|_{L^1}$ converges to zero for $T, j \to \infty$, we obtain from Corollary 5 the convergence of the second expected value:

$$
\begin{aligned}
&\mathbb{E}_{a_0+d_\varepsilon+d_{\varepsilon'}}\left[\Lambda_r(X^{(a_\varepsilon)}, X^{(a_0)})^2 \Lambda_r(X^{(a_{\varepsilon'})}, X^{(a_0+d_\varepsilon+d_{\varepsilon'})})^2\right] \\
&\qquad \leq \mathbb{E}_{a_0}\left[\Lambda_r(X^{(a_\varepsilon)}, X^{(a_0)})^4 \Lambda_r(X^{(a_0+d_\varepsilon+d_{\varepsilon'})}, X^{(a_0)})\right]^{1/2} \bullet \\
&\qquad\quad \bullet \mathbb{E}_{a_0+d_\varepsilon+d_{\varepsilon'}}\left[\Lambda_r(X^{(a_{\varepsilon'})}, X^{(a_0+d_\varepsilon+d_{\varepsilon'})})^4\right]^{1/2} \\
&\qquad \leq E_{a_0}\left[\Lambda_r(X^{(a_\varepsilon)}, X^{(a_0)})^8\right]^{1/4} \mathbb{E}_{a_0}\left[\Lambda_r(X^{(a_0+d_\varepsilon+d_{\varepsilon'})}, X^{(a_0)})^2\right]^{1/4} \bullet \\
&\qquad\quad \bullet \mathbb{E}_{a_0+d_\varepsilon+d_{\varepsilon'}}\left[\Lambda_r(X^{(a_0+d_{\varepsilon'})}, X^{(a_0+d_\varepsilon+d_{\varepsilon'})})^4\right]^{1/2} \\
&\qquad \to 1.
\end{aligned}
$$

By stationarity it thus suffices to bound

$$
\begin{aligned}
&\mathbb{E}_{a_0+d_\varepsilon+d_{\varepsilon'}}\left[\exp\left(2\int_r^T \left(\int_{-r}^0 X(t+s) d_\varepsilon(s)\, ds\right)\left(\int_{-r}^0 X(t+s) d_{\varepsilon'}(s)\, ds\right) dt\right)\right]^{1/2} \\
&= \mathbb{E}_{a_0+d_\varepsilon+d_{\varepsilon'}}\left[\exp\big(2\langle Q_{T-r} d_\varepsilon, d_{\varepsilon'}\rangle\big)\right]^{1/2}. \qquad (8.3.6)
\end{aligned}
$$

Inspired from the calculations for the model problem, we split the expected value into three factors

$$
\begin{aligned}
\mathbb{E}_{a_0+d_\varepsilon+d_{\varepsilon'}}\left[\exp\big(2\langle Q_T d_\varepsilon, d_{\varepsilon'}\rangle\big)\right] &= \mathbb{E}_{a_0+d_\varepsilon+d_{\varepsilon'}}\left[\exp\big(2\langle (Q_T - TQ_{a_0+d_\varepsilon+d_{\varepsilon'}}) d_\varepsilon, d_{\varepsilon'}\rangle\big)\right] \\
&\quad \bullet \exp\big(2T\langle (Q_{a_0+d_\varepsilon+d_{\varepsilon'}} - Q_W) d_\varepsilon, d_{\varepsilon'}\rangle\big) \\
&\quad \bullet \exp\big(2T\langle Q_W d_\varepsilon, d_{\varepsilon'}\rangle\big) \qquad (8.3.7)
\end{aligned}
$$

and show that the first two factors converge to one uniformly over $\varepsilon, \varepsilon'$ in the limit $T \to \infty$.

We want to apply Proposition 6 with $\mu := \frac{d_\varepsilon}{\|d_\varepsilon\|_{L^1}}$, which is possible for the asymptotically vanishing values of α:

$$\alpha := 2|R_j|T^{1/2}\gamma 2^{-3j/2}\|d_\varepsilon\|_{L^1} \sim T^{1/2}\gamma^2 \sim \gamma_0^2 T^{\frac{-s+0.25}{2s+2.5}} \to 0, \quad s > \frac{1}{4}.$$

Hence, by Hölder's inequality we derive the bound for large values of T

$$\begin{aligned}
&\mathbb{E}_{a_0+d_\varepsilon+d_{\varepsilon'}}\left[\exp\left(2\langle (Q_T - TQ_{a_0+d_\varepsilon+d_{\varepsilon'}})d_\varepsilon, d_{\varepsilon'}\rangle\right)\right] \\
&\le \prod_{k\in R_j} \mathbb{E}_{a_0+d_\varepsilon+d_{\varepsilon'}}\left[\exp\left(2|R_j|\langle (Q_T - TQ_{a_0+d_\varepsilon+d_{\varepsilon'}})d_\varepsilon, \gamma\varepsilon'_k\psi_{jk}\rangle\right)\right]^{1/|R_j|} \\
&\le \frac{1}{1 - C_1\gamma^4 T} \\
&\le \frac{1}{1 - C_2\gamma_0^4 T^{\frac{-s+0.25}{2s+2.5}}}.
\end{aligned}$$

Again, by the uniformity of the estimate in Proposition 6 we can choose a universal constant C_1 for all $a \in \mathcal{K}_\rho \subset M(S + \|a_0\|_{TV}, \delta)$, which implies that the first factor of the right hand side in (8.3.7) converges uniformly to one for $T \to \infty$.

The covariance function q_a, when restricted to $[-r, 0]$, lies in $H^3([-r, 0])$ by Proposition 3. We claim that then $h(t) := q_a(t) + \frac{1}{2}|t|$, $t \in [-r, r]$, even lies in $H^3([-r, r])$. For $t \in [0, r]$ we obtain

$$h(t) = q_a(0) + (q'_a(0+) + \tfrac{1}{2})t + \tfrac{1}{2}q''_a(0+)t^2 + \int_0^t \int_0^u \int_0^v q'''_a(w)\,dw\,dv\,du,$$

which due to $q'_a(0+) = -\frac{1}{2}$ also holds for $t \in [-r, 0]$ if the symmetry relations $q''_a(0+) = q''(0-)$ and $q'''_a(v) = -q'''(-v)$ are taken into account. Therefore q''' is the third weak derivative of h and lies in $L^2([-r, r])$, whence $h \in H^3([-r, r])$ follows.

We obtain for $a \in \mathcal{W}^0_{2,2}$ due to $\int \psi = 0$, the Cauchy-Schwarz-inequality and the characterisation of H^3

$$\begin{aligned}
|\langle (Q_a - Q_W)d_\varepsilon, d_{\varepsilon'}\rangle| &= \left|\int_{-r}^0 \int_{-r}^0 (q_a(u-v) - \min(u,v) - r - 1)d_\varepsilon(u)d_{\varepsilon'}(v)\,du\,dv\right| \\
&= \left|\int_{-r}^0 \int_{-r}^0 (h(u-v) - r - \tfrac{u+v}{2})d_\varepsilon(u)d_{\varepsilon'}(v)\,du\,dv\right| \\
&\le \gamma^2 \sum_{k,k'\in R_j} \left|\int_{-r}^0 \int_{-r}^0 h(u-v)\psi_{jk}(u)\psi_{jk'}(v)\,du\,dv\right| \\
&\lesssim \gamma^2 (2^{2j})^{1/2} \left(\sup_{v\in[-r,0]} \sum_{k,k'\in R_j} \langle h(\bullet - v), \psi_{jk}\rangle^2 \|\psi_{jk'}\|_{L^1}^2\right)^{1/2} \\
&\lesssim \gamma^2 2^j \left(\sup_{v\in[-r,0]} 2^{-6j} \sum_{k\in R_j} 2^{6j}\langle h(\bullet - v), \psi_{jk}\rangle^2\right)^{1/2} \\
&\lesssim \gamma^2 2^j 2^{-3j}\|h\|_{H^3([-r,r])} \\
&\lesssim T^{-\frac{2s+3}{2s+2.5}}(\|q_a\|_{H^3([0,r])} + 1).
\end{aligned}$$

With a look at Proposition 4 this calculation for $a = a_0 + d_\varepsilon + d_{\varepsilon'}$ shows that the second factor in (8.3.7) also converges uniformly to one for $T \to \infty$. We have thus reduced the SDDE case to the model problem case. We conclude that the bound $\exp(C_6\gamma_0^4)(1+o(T))$ holds for the expected value in (8.3.5) with $o(T) \to 0$ for $T \to \infty$. This finally proves as for the model problem the lower bound by choosing γ_0 and thus C_η sufficiently small. □

Remarks 14.

- *As a consequence of the theorem we obtain that for* $\rho(T)T^{\frac{s}{2s+2.5}} \to 0$ *the sum of the error probabilities tends to one. This means that we cannot distinguish between hypothesis and alternative, because this bound is attained by the constant decision functions* $\Phi = 0$ *and* $\Phi = 1$.
- *The restriction* $s > \frac{1}{4}$ *is probably only of technical nature. We have used a very rough Hölder inequality estimate in order to apply Proposition 6. A more subtle analysis could show that this restriction is not necessary.*
- *The lower bound and its proof remain the same when we allow for randomized decision functions* Φ *taking values in* $[0,1]$ *and consider* $\mathbb{E}_{a_0}[\Phi] + \mathbb{E}_a[1-\Phi]$ *instead of* $\mathbb{P}_{a_0}(\Phi = 1) + \mathbb{P}_a(\Phi = 0)$.

Chapter 9

Conclusion

The problem of nonparametric inference for the weight function is now fairly well understood under stationarity assumptions. In particular, its treatment as an ill-posed inverse problem has been shown to be adequate by deriving minimax rates for the estimation and the testing problem. Even adaptive estimation methods work in a rate-optimal way. In principle, all the techniques developed for a signal plus noise model can be transferred to the SDDE model by regarding $\frac{1}{T}b_T$ as noisy observation of $Q_a a$. For instance, the detection of change points should not pose any problems, because the adaptive wavelet thresholding algorithm has been shown to work, and could for instance yield an estimator of the memory length r. From a practitioner's point of view the methods developed here could at least serve to get an intuition of "what is going on" before using more specific tools, e.g. parametric methods. Another important application is the testing against a nonparametric alternative, which for parametric model assumptions should be obligatory.

Some minor details that have not been settled concern for example the restrictions $s > \frac{1}{2}$ in Theorem 4 and $s > \frac{1}{4}$ in Theorem 10 or the correct power of the logarithmic factor in Corollary 11. More consequences for a general theoretical understanding are to be expected from results concerning the L^1-boundedness of the estimator for the misspecified model in Section 5.5 or concerning the injectivity of the operators Q_T and $\widehat{Q}_T$. Another point that needs further study is the case of discrete time observations; especially lower bounds for $\Delta \to 0$ and the problem of identifiability for constant Δ should be studied more intensively. Also an extension of the test model as discussed in Remark 10 would be advisable.

The strong relation between the SDDE case and the Gaussian model problem suggests an investigation whether this closeness of the models can be made precise in Le Cam's sense of asymptotic equivalence of statistical experiments. It would yield the mathematical justification of the statement made above that everything that is known for the model problem also works in the SDDE case and would imply all the results obtained here almost "for free". However, some of the estimates derived here will certainly be needed again for the proof of this asymptotic equivalence. Besides these theoretical questions an implementation of the adaptive estimator and the hypothesis test, simulation studies of their accuracy and applications to real data are also of considerable interest.

The model assumptions treated here are quite specific. Since the Galerkin estimator is close to the maximum likelihood estimator for a general affine SDDE, it should still yield reasonable results for weight functions that do not admit stationary solutions. Note however that in this case the trajectory will be growing exponentially and the observation and treatment of the data become problematic. The same methods of inference can also be applied to multi-dimensional affine SDDEs where the theory is comparable, but the weights are in this case matrices of measures or functions. Thus, the curse of dimensionality will apply with a term that is quadratic in the dimension and the problem becomes highly ill-posed, unless certain submodel assumptions are made. Another line of research consists of keeping the Gaussian model, but using fractional Brownian motion as driving term. The covariance operator of fractional Brownian motion maps L^2 to some Sobolev space H^α corresponding to its index and it would be interesting to see whether this change in the degree of ill-posedness results in a different estimation rate. Before that, a thorough theoretical study will be necessary because fractional Brownian motion is in general not a semimartingale.

The continuous time limit of general ARMA time series yields a memory effect in the diffusion term which might read like $\sigma_1 dW(t-1) + \sigma_0 dW(t)$. A stochastic version of so-called neutral delay equations is obtained by equations with a drift of the form $(\int_{t-r}^{t} g(s-t)dX(s))dt$ with $g \in L^2([-r,0])$ and with additive white noise (cf. also equation (3.2.1)). By partial integration, neglecting boundary terms, this is a generalisation of affine SDDEs. The latter two equations are inhomogeneous linear SDDEs that have apparently not been well studied yet, but should have many features in common with affine SDDEs. Further variations of the drift or diffusion term could be thought of, but abandoning "Gaussianity" will require more refined techniques, because our methods very much rely on powerful results for infinite-dimensional Gaussian measures. So let us close with the master's voice:

> *There are problems to whose solution I would attach an infinitely greater importance than to those of mathematics, for example touching ethics, or our relation to God, or concerning our destiny and our future; but their solution lies wholly beyond us and completely outside the province of science.*
>
> Johann Carl Friedrich Gauß (1777-1855)
>
> (from J. R. Newman, *The World of Mathematics*)

Appendix A

Function spaces and wavelets

In this appendix results on function spaces, especially Sobolev and Besov spaces, and their characterisation by wavelet bases are presented. A thorough introduction into these topics is provided for instance by Triebel (1983) (only function spaces), Meyer (1995), Härdle et al. (1998) or Cohen (2000). We start with the definition of L^2-Sobolev spaces H^s of regularity s and state their spectral characterisation and the Sobolev embedding theorem. Besov spaces $B^s_{p,q}$ are defined in terms of the modulus of continuity, the embedding properties are stated and illustrated by examples. Finally, the concept of wavelet bases is introduced. The characterisation of Besov spaces by regular wavelet bases and two subsequently useful corollaries close the appendix.

A.1 Sobolev and Hölder-type spaces

Let us introduce the scale of Sobolev spaces $H^s(\mathbb{R})$, $s \geq 0$, which combine the advantages of a differentiability description and of a Hilbert space structure. For $s = m \in \mathbb{N}_0$ set

$$H^m(\mathbb{R}) := \{f \in L^2(\mathbb{R}) \mid f^{(i)} \in L^2(\mathbb{R}) \text{ for all } i = 0, \dots, m\},$$

where $f^{(i)}$ denotes the i-th derivative of f in a weak (distributional) sense. These spaces are Hilbert spaces with respect to the following norm and scalar product

$$\|f\|_m^2 := \sum_{i=0}^m \|f^{(i)}\|_{L^2}^2, \quad \langle f, g \rangle_m := \sum_{i=0}^m \langle f^{(i)}, g^{(i)} \rangle_{L^2}.$$

By properties of the Fourier transform we obtain

$$\|f\|_m^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} (1 + \xi^2)^m |\hat{f}(\xi)|^2 \, d\xi.$$

This spectral representation provides a method for defining Sobolev spaces of non-integer regularity $s \geq 0$:

$$H^s(\mathbb{R}) := \{f \in L^2(\mathbb{R}) \mid \|f\|_s^2 := \int_{-\infty}^{\infty} (1 + \xi^2)^s |\hat{f}(\xi)|^2 \, d\xi < \infty\}.$$

These spaces are Hilbert spaces and coincide with those constructed by the abstract complex interpolation method.

The spaces $H^m(I)$, I an interval, are either derived from $H^m(\mathbb{R})$ by restriction or equivalently by the same inner definition. Note that the Fourier transform method cannot be used. For non-integer $s \geq 0$ an abstract interpolation method or the restriction method can be applied. For $s < 0$ the spaces H^s are defined by duality as the space of all distributions F satisfying

$$\|F\|_s := \sup_{\substack{h \in H^{-s} \\ \|h\|_{-s}=1}} \langle F, h \rangle < \infty.$$

As an example consider the function $f := \mathbf{1}_{[-1,1]}$. Its Fourier transform is $\hat{f}(\xi) = \sin(\xi)\xi^{-1}$ and by the spectral definition we infer that f lies in $H^s(\mathbb{R})$ for all $s < \frac{1}{2}$. It is in $H^s([-1,1])$ for all $s \geq 0$.

For $H^s(I)$, $I \subset \mathbb{R}$ a bounded or unbounded interval, the following embedding relations hold:

$$H^s \subset H^{s'} \text{ for all } s' < s, \tag{A.1.1}$$

$$H^s \subset C^\alpha \text{ for all } \alpha < s - \tfrac{1}{2}, \tag{A.1.2}$$

where C^α, $\alpha = m + \beta$, $m \in \mathbb{N}_0$, $\beta \in [0,1)$, is the Banach space

$$C^\alpha(I) := \{f \in C(I) \mid f^{(i)} \text{ is bounded, } i = 0, \dots, m;\ f^{(m)} \text{ is } \beta\text{-Hölder continuous}\}$$

with norm

$$\|f\|_{C^\alpha} := \sum_{i=0}^m \|f^{(i)}\|_\infty + \sup_{\substack{x,h \in \mathbb{R} \\ h \neq 0}} \frac{|f^{(m)}(x+h) - f^{(m)}(x)|}{|h|^\beta}.$$

Later on we shall also need the Lipschitz-type spaces for $m \in \mathbb{N}_0$

$$C^{m,1}(I) := \{f \in C(I) \mid f^{(i)} \text{ is bounded, } i = 0, \dots, m;\ f^{(m)} \text{ is Lipschitz-continuous}\}$$

with norm

$$\|f\|_{C^{m,1}} := \sum_{i=0}^m \|f^{(i)}\|_\infty + \sup_{x,h \in \mathbb{R}} \frac{|f^{(m)}(x+h) - f^{(m)}(x)|}{|h|}.$$

Note that throughout this work the derivative at the boundary of I is taken as the one-sided derivative and sometimes the notation $f'(a+)$ is used for the derivative of f to the right at a.

In particular the norm estimate $\|\bullet\|_{C^\alpha} \lesssim \|\bullet\|_{H^s}$ holds for $\alpha < s - \frac{1}{2}$. In the case of a bounded interval I, the embeddings are compact, i.e. an H^s-bounded set is relatively compact in $H^{s'}$-norm and in C^α-norm for s' and α as above.

More generally, one defines for any open set $O \subset \mathbb{R}^d$ (or its closure $\overline{O}$) with Euclidean norm $|\bullet|$ and for $m \in \mathbb{N}_0$, $p \in [1, \infty]$ the L^p-Sobolev spaces

$$W^{m,p}(\overline{O}) = W^{m,p}(O) := \{f \in L^p(O) \mid f^{(\alpha)} \in L^p(O),\ |\alpha| \leq m\},$$

where $\alpha = (\alpha_1, \dots, \alpha_d)$ denotes a multi-index with $|\alpha| = \sum_i \alpha_i$. With respect to the norm

$$\|f\|_{W^{m,p}} := \sum_{0 \leq |\alpha| \leq m} \|f^{(\alpha)}\|_{L^p},$$

these spaces are Banach spaces and by interpolation techniques one can define $W^{s,p}(O)$ for all $s \geq 0$. For $s \in (0,1)$, $p < \infty$, the spaces $W^{s,p}(O)$ have the following equivalent norm

$$\|f\|_{W^{s,p}(O)} \sim \|f\|_{L^p(O)} + \left(\iint_{O \times O} \frac{|f(x) - f(y)|^p}{|x-y|^{sp+d}} \, dx \, dy \right)^{1/p}.$$

The general Sobolev embedding theorem states $W^{s,p}(O) \subset C^\alpha(O)$ for all $\alpha < s - \frac{d}{p}$ with continuous embeddings.

A.2 Besov spaces

An even larger scale of function spaces is given by the Besov spaces $B^s_{p,\alpha}$, measuring the regularity s in an L^p-sense with an additional fine-tuning parameter $\alpha \in [1, \infty]$. These spaces appear naturally, e.g. from the real interpolation method or in nonlinear approximation theory. We shall however present the classical definition via the modulus of continuity, which requires the least amount of technical preparations.

Definition 14. *Let $I \subset \mathbb{R}$ be a bounded or unbounded interval, $\Delta_h f(x) := f(x+h) - f(x)$ and $I_h := \{x \in I \mid x \pm h \in I\}$. Then the n-th order L^p-modulus of smoothness is defined by*

$$\omega_n(f, \varepsilon)_p := \sup_{|h| \le \varepsilon} \|\Delta_h^n f\|_{L^p(I_{nh})},$$

where Δ_h^n denoting the n-fold application of Δ_h. For $p, \alpha \in [1, \infty]$ and $s > 0$ set

$$\|f\|_{s,p,\alpha} := \|f\|_{L^p(I)} + \left(\int_0^1 \left(\frac{\omega_n(f,t)_p}{t^s} \right)^\alpha \frac{dt}{t} \right)^{1/\alpha}$$

with the usual modification $\sup_t \omega_n(f,t)_p t^{-s}$ for $\alpha = \infty$ and with $n = \lfloor s \rfloor + 1$. The Besov space *$B^s_{p,\alpha}(I)$ is the subspace of $L^p(I)$*

$$B^s_{p,\alpha}(I) := \{f \in L^p(I) \mid \|f\|_{s,p,\alpha} < \infty\},$$

which is a Banach space with respect to the norm $\|\bullet\|_{B^s_{p,\alpha}} := \|\bullet\|_{s,p,\alpha}$. On a bounded interval I an equivalent norm is given by (n as above)

$$\|f\|_{B^s_{p,\alpha}} \sim \|f\|_{L^p} + \|f^{(n-1)}\|_{s-(n-1),p,\alpha}.$$

For $s = 0$ or even more general parameter values $s \in \mathbb{R}$, $p, \alpha > 0$ the spaces $B^s_{p,\alpha}$ can still be defined using the Littlewood-Paley decomposition or wavelets. We send the reader to Triebel (1983, Def. 2.3.1) or Meyer (1995, Sec. 2.9) for its exact definition.

How do Besov spaces relate to Sobolev and Hölder-type spaces? This and a corresponding embedding theorem are the content of the next proposition, which collects statements from Triebel (1983).

Proposition 16. *For all $s \in \mathbb{R}$ the identity $B^s_{2,2} = H^s$ holds with equivalent norms. For non-integer $s > 0$ we have $B^s_{p,p} = W^{s,p}$ and $B^s_{\infty,\infty} = C^s$. For $m \in \mathbb{N}_0$ we only have $B^m_{p,p\wedge 2} \subset W^{m,p} \subset B^m_{p,p\vee 2}$ and $B^m_{\infty,1} \subset C^m \subset C^{m,1} \subset B^m_{\infty,\infty}$.*

The fact that the parameter α only plays a minor role is reflected by the first of the following embedding relations

- $B^s_{p,\alpha} \subset B^{s'}_{p,\alpha'}$, *$s > s'$, any α, α';*
- $B^s_{p,\alpha} \subset B^s_{p',\alpha}$, *$p > p'$;*
- $B^s_{p,\alpha} \subset B^s_{p,\alpha'}$, *$\alpha < \alpha'$;*
- *even sharper, the Sobolev embedding theorem generalizes to*

$$B^s_{p,\alpha} \subset B^{s'}_{p',\alpha} \text{ for } s \ge s' \text{ and } s - \tfrac{1}{p} \ge s' - \tfrac{1}{p'}; \tag{A.2.3}$$

as a special case $B^s_{p,\alpha} \subset C^{s'}$ for $s - \frac{1}{p} > s'$ follows.

The first embedding is compact for Besov spaces on bounded intervals.

An illustration of the role played by the different parameters is given by the result that paths of Brownian motion almost surely lie in $B^{1/2}_{p,\infty}$ for all $p \in [1, \infty)$, but, as is well-known, neither in $C^{1/2}$ nor in $H^{1/2}$. As another example consider again $f = \mathbf{1}_{[-1,1]}$. It satisfies $\omega_1(f, \varepsilon)_p = (2\varepsilon)^{1/p}$ and therefore $\|f\|_{1,1,\infty} < \infty$. By embedding or directly we see that $f \in B^s_{p,\infty}$ holds for all $s \le \frac{1}{p}$ (cf. $f \in H^s$, $s < \frac{1}{2}$, obtained previously). It is a good example for the phenomenon that by some loss of integrability we obtain higher regularity. This is the essential idea used in the theory of nonlinear approximation and adaptive estimation.

A.3 Wavelets

We immediately start with the definition of an orthonormal wavelet basis in $L^2(\mathbb{R})$.

Definition 15. *For $j \in \mathbb{Z}$ and $k \in \mathbb{Z}$ introduce the multi-index $\lambda = (j, k)$ and put $|\lambda| := |(j, k)| := j$. A* wavelet basis *$(\psi_\lambda)_\lambda$ is an orthonormal basis of functions in $L^2(\mathbb{R})$, derived from one function $\psi \in L^2(\mathbb{R})$ by translations and dilations*

$$\psi_\lambda(x) := \psi_{jk}(x) := 2^{j/2}\psi(2^j x - k).$$

Furthermore set V_j as the closure of $\operatorname{span}(\psi_\lambda,\ |\lambda| \le j)$. *By $P_j : L^2([-r, 0]) \to V_j$ we denote the orthogonal projection onto V_j.*

An example of a function ψ yielding such a wavelet basis is the Haar-function $\psi = \mathbf{1}_{[0,\frac{1}{2}]} - \mathbf{1}_{[\frac{1}{2},1]}$. In this case V_j is the space of all functions in $L^2(\mathbb{R})$ that are modulo Lebesgue null sets constant on the intervals $[2^{-j}k, 2^{-j}(k+1))$, $k \in \mathbb{Z}$. The existence of wavelet bases besides the Haar-wavelet and related spline wavelets is a nontrivial fact, in particular compactly supported wavelets of arbitrary high regularity exist, the so-called Daubechies-wavelets Meyer (1995, Section 3.8).

Cohen et al. (1993) constructed orthonormal wavelet bases on a bounded interval I. The basis functions are obtained by restricting the Daubechies-wavelets to this interval. Wavelet functions ψ_λ whose support crosses the boundary of I are suitably corrected in order to keep the orthogonality and approximation properties. These corrected functions are still denoted by ψ_λ even if they are not directly derived from ψ. A consequence of this construction is that only multi-indices $\lambda = (j, k)$ with $|k| \lesssim 2^j$ are used and that the spaces V_j are finite-dimensional, whence we can start off with a space V_{-1} and an orthonormal basis $(\psi_{-1,k})_k$ of V_{-1}. Then any function $f \in L^2(I)$ has the wavelet decomposition

$$f = \sum_\lambda \langle f, \psi_\lambda\rangle \psi_\lambda = \sum_{j \ge -1} \sum_k \langle f, \psi_{jk}\rangle \psi_{jk}$$

and its approximation $P_{j_0} f$ on the resolution level $j_0 \in \mathbb{N}$ is given by

$$P_{j_0} f = \sum_{|\lambda| \le j_0} \langle f, \psi_\lambda\rangle \psi_\lambda = \sum_{j=-1}^{j_0} \sum_k \langle f, \psi_{jk}\rangle \psi_{jk},$$

where the second sum is taken over all k with $\operatorname{supp}(\psi_{jk}) \cap I \ne \varnothing$. Note that summation over $|\lambda| \le j_0$ in this work will always mean summation over (j, k) for all $j \le j_0$ and all corresponding values of k

Wavelets are like tailor-made for the description of Besov spaces.

Definition 16. *A wavelet basis (ψ_λ) will be called* s-regular *on the bounded or unbounded interval I if the following two conditions are satisfied:*

1. *For all $\sigma \in [-s,s]$, $p,\alpha \in [1,\infty]$ the function or distribution f is an element of $B^{\sigma}_{p,\alpha}(I)$ if and only if*

$$\|P_{-1}f\|_{L^p} + \left(\sum_{j=0}^{\infty} 2^{\alpha j(\sigma+\frac{1}{2}-\frac{1}{p})}\left(\sum_k |\langle f,\psi_{jk}\rangle|^p\right)^{\alpha/p}\right)^{1/\alpha} < \infty.$$

   The above expression constitutes a norm equivalent to $\|\bullet\|_{\sigma,p,\alpha}$.

2. *For all $k = 0,\dots,\lfloor s\rfloor$ the vanishing moment property is fulfilled*

$$\int_{-\infty}^{\infty} x^k \psi(x)\,dx = 0.$$

In particular, an s-regular wavelet basis (ψ_λ) of $L^2([-r,0])$ yields for $f \in H^s([-r,0])$

$$\|f\|^2_{H^s} \sim \sum_{\lambda} 2^{2s|\lambda|}\langle f,\psi_\lambda\rangle^2 \quad \text{and} \quad \|P_{j_0}f\|^2_{H^s} \sim \sum_{|\lambda|\le j_0} 2^{2s|\lambda|}\langle f,\psi_\lambda\rangle^2.$$

From Meyer (1995, Prop. 2.9.4, Thm. 2.6.4) and Cohen et al. (1993) we immediately obtain (only mind the different notion of s-regularity there):

Theorem 11. *s-regular wavelet bases exist for any $s > 0$. Moreover, they may be chosen to have compact support.*

Observe that the Besov space embedding properties are simple consequences of these characterisations by Hölder's inequality. Two short corollaries are of central importance in our context.

Corollary 12. *The spaces (V_j) are s-approximating in the sense of Definition 7 for $n \sim 2^j$ if the wavelet basis is $(s \vee 1)$-regular.*

Proof. The Jackson-inequality follows easily:

$$\|f - P_j f\|^2_{L^2} = \sum_{|\lambda|>j} \langle f,\psi_\lambda\rangle^2 \le 2^{-2js} \sum_{|\lambda|>j} 2^{2|\lambda|s}\langle f,\psi_\lambda\rangle^2 \lesssim 2^{-2js}\|f\|^2_{H^s}.$$

For the establishment of the Bernstein inequality we use Proposition 9. The first of its conditions is satisfied, because for $f \in H^1$

$$\|f - P_j f\|^2_{H^1} \sim \sum_{|\lambda|>j} 2^{2|\lambda|}\langle f,\psi_\lambda\rangle^2 \to 0 \text{ as } j \to \infty.$$

The second condition is fulfilled due to

$$\|v_j\|^2_{H^1} \sim \sum_{|\lambda|\le j} 2^{2|\lambda|}\langle v_j,\psi_\lambda\rangle^2 \le 2^{2j}\|v_j\|^2_{L^2}$$

uniformly for all $v_j \in V_j$. □

Remark 11. *Even more generally, for $f \in H^{s+\rho}$ and an $(s+\rho)$-regular wavelet basis, $s,\rho \ge 0$, we obtain the Jackson inequality in H^ρ*

$$\|f - P_j f\|_{H^\rho} \lesssim \left(2^{-2js}\sum_{|\lambda|>j} 2^{2(\rho+s)|\lambda|}\langle f,\psi_\lambda\rangle^2\right)^{1/2} \lesssim 2^{-js}\|f\|_{H^{s+\rho}}.$$

Corollary 13. *Let $f \in B^s_{p,\infty}([-r,r])$ be given with $s \geq 0$, $p \in [1,\infty]$ and suppose that (ψ_λ) is a compactly supported s-regular wavelet basis of $L^2([-r,0])$. Then*

$$\left|\int_{-r}^{0}\int_{-r}^{0} f(x-y)\psi_\lambda(x)\psi_\lambda(y)\,dy\,dx\right| \leq C\|f\|_{B^s_{p,\infty}([-r,r])}2^{-|\lambda|(s+1-\frac{1}{p})}$$

holds with a constant $C > 0$ independent of f and of the multi-index λ. In particular, for $f \in C^{m,1}([-r,r])$, $m \in \mathbb{N}_0$, we obtain the bound $C\|f\|_{C^{m,1}}2^{-|\lambda|(m+2)}$.

Proof. Note that for $y \in [-r,0]$ the function $f(\bullet - y)|_{[-r,0]}$ lies in $B^s_{p,\infty}$. By the s-regularity of (ψ_λ) we find

$$\begin{aligned}
&\left|\int_{-r}^{0}\int_{-r}^{0} f(x-y)\psi_\lambda(x)\psi_\lambda(y)\,dy\,dx\right| \\
&\qquad \leq \sup_{y\in[-r,0]} |\langle f(\bullet - y), \psi_\lambda\rangle| \|\psi_\lambda\|_{L^1} \\
&\qquad \lesssim \sup_{y\in[-r,0]} \|f(\bullet - y)\|_{B^s_{p,\infty}} 2^{-|\lambda|(s+\frac{1}{2}-\frac{1}{p})}2^{-|\lambda|/2}\|\psi\|_{L^1} \\
&\qquad \leq 2^{-|\lambda|(s+1-\frac{1}{p})}\|f\|_{B^s_{p,\infty}([-r,r])}\|\psi\|_{L^1}.
\end{aligned}$$

The constant involved is only due to the Besov norm description by the wavelet basis. The continuous embedding $C^{m,1} \subset B^{m+1}_{\infty,\infty}$ yields the last statement. □

Bibliography

Ahlfors, Lars V. (1979). *Complex analysis. An introduction to the theory of analytic functions of one complex variable. 2nd ed.* McGraw-Hill, Kogakusha.

Baumeister, Johann (1987). *Stable solution of inverse problems.* Vieweg, Braunschweig/Wiesbaden.

Bellman, Richard and Cooke, Kenneth L. (1963). *Differential-difference equations.* Academic Press, New York.

Bennett, Colin and Sharpley, Robert (1988). *Interpolation of operators.* Pure and Applied Mathematics, Vol. 129., Academic Press, Boston.

Billingsley, Patrick (1968). *Convergence of probability measures. 1st ed.* Wiley, Chichester.

Brockwell, Peter J. and Davis, Richard A. (1996). *Introduction to time series and forecasting.* Springer, New York.

Cohen, Albert (2000). Wavelet methods in numerical analysis. In Ciarlet, P. G., editor, *Handbook of numerical analysis. Vol. 7.* North-Holland/ Elsevier, Amsterdam.

Cohen, Albert and Daubechies, Ingrid and Vial, Pierre (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.*, 1(1):54–81.

Da Prato, Giuseppe and Zabczyk, Jerzy (1992). *Stochastic equations in infinite dimensions.* Cambridge University Press, Cambridge.

Dehling, Herold and Philipp, Walter (1982). Almost sure invariance principles for weakly dependent vector-valued random variables. *Ann. Probab.*, 10:689–701.

Diekmann, Odo and van Gils, Stephan A. and Verduyn Lunel, Sjoerd M. and Walther, Hans-Otto (1995). *Delay equations. Functional-, complex-, and nonlinear analysis.* Springer, New York.

Donoho, David L. (1995). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Comput. Harmon. Anal.*, 2(2):101–126.

Dunford, N. and Schwartz, J.T. (1963). *Linear operators. Part II: Spectral theory. Self adjoint operators in Hilbert space.* Wiley-Interscience, New York.

Elsanosi, Ismail and Øksendal, Bernt and Sulem, Agnès (2000). Some solvable stochastic control problems with delay. Preprint, University of Oslo.

Gushchin, Alexander A. and Küchler, Uwe (1999). Asymptotic inference for a linear stochastic differential equation with time delay. *Bernoulli*, 5(6):1059–1098.

Gushchin, Alexander A. and Küchler, Uwe (2000). On stationary solutions of delay differential equations driven by a Lévy process. *Stochastic Proc. Appl.*, 88:195–211.

Hale, Jack K. and Verduyn Lunel, Sjoerd M. (1993). *Introduction to functional differential equations.* Springer, New York.

Hämarik, Uno (1983). Projection methods for regularizing ill-posed problems (in Russian). *Tr. Vychisl. Tsentra Tartusk*, 50:69–90.

Härdle, Wolfgang and Kerkyacharian, Gerard and Picard, Dominique and Tsybakov, Alexander (1998). *Wavelets, approximation, and statistical applications.* Springer, Berlin.

Hitsuda, M. (1968). Representation of Gaussian processes equivalent to Wiener process. *Osaka J. Math.*, 5:299–312.

Hoffmann, Marc (1999). Adaptive estimation in diffusion processes. *Stochastic Processes Appl.*, 79(1):135–163.

Horowitz, Joel and Spokoiny, Vladimir (1999). An adaptive, rate-optimal test of a parametric model against a nonparametric alternative. Preprint 542, Weierstraß-Institut Berlin (WIAS).

Hu, Yaozhong and Mohammed, Salah-Eldin and Yan, Feng (2001). Numerical solutions of stochastic differential systems with memory. Preprint, Southern Illinois University, Carbondale.

Ibragimov, I.A. and Khas'minskii, R.Z. (1981). *Statistical estimation. Asymptotic theory.* Springer, New York.

Ibragimov, I.A. and Rozanov, Y.A. (1978). *Gaussian random processes. Translated by A. B. Aries.* Springer, New York.

Ingster, Yu.I. (1993a). Asymptotically minimax hypothesis testing for nonparametric alternatives. I. *Math. Methods Stat.*, 2(2):85–114.

Ingster, Yu.I. (1993b). Asymptotically minimax hypothesis testing for nonparametric alternatives. II. *Math. Methods Stat.*, 2(3):171–189.

Ingster, Yu.I. (1993c). Asymptotically minimax hypothesis testing for nonparametric alternatives. III. *Math. Methods Stat.*, 2(4):249–268.

Itô, K. and Nisio, M. (1964). On stationary solutions of a stochastic differential equation. *J. Math. Kyoto Univ.*, 4:1–75.

Karatzas, Ioannis and Shreve, Steven E. (1991). *Brownian motion and stochastic calculus. 2nd ed.* Springer, New York.

Katznelson, Yitzhak (1976). *An introduction to harmonic analysis. 2nd corr. ed.* Dover Publications, New York.

Kerkyacharian, Gérard and Picard, Dominique (2000). Thresholding algorithms, maxisets and well-concentrated bases. *Test*, 9(2):283–344.

Kolmanovskii, V. and Myshkis, A. (1992). *Applied theory of functional differential equations.* Mathematics and Its Applications. Soviet Series. 85. Kluwer, Dordrecht.

Küchler, Uwe and Kutoyants, Yuri A. (2000). Delay estimation for some stationary diffusion-type processes. *Scand. J. Statist.*, 27:1–10.

Küchler, Uwe and Mensch, Beatrice (1992). Langevin's stochastic differential equation extended by a time-delayed term. *Stochastics Stochastics Rep.*, 40(1/2):23–42.

Küchler, Uwe and Sørensen, Michael (1997). *Exponential families of stochastic processes.* Springer, New York.

Kutoyants, Yury and Mourid, Tahar (1994). Estimation par la distance minimale pour un processus de type diffusion avec retards. (Minimum distance estimation for a diffusion-type process with delays). *Publ. Inst. Stat. Univ. Paris*, 38(2):3–18.

Kutoyants, Yu.A. and Mourid, T. and Bosq, D. (1992). Estimation parametrique d'un processus de diffusion avec retards. (Parametric estimation of a diffusion process with delays). *Ann. Inst. Henri Poincare, Probab. Stat.*, 28(1):95–106.

Liptser, Robert S. and Shiryaev, Albert N. (2001). *Statistics of random processes. 1: General theory. 2nd ed.* Springer, Berlin.

Mao, Xuerong (1997). *Stochastic differential equations and their applications.* Horwood Publishing, Chichester.

Mensch, Beatrice (1989). *Über einige lineare stochastische Funktional-Differentialgleichungen: Stationarität und statistische Aspekte.* PhD thesis, Humboldt Universität zu Berlin.

Meyer, Yves (1995). *Wavelets and operators.* Cambridge University Press, Cambridge.

Mohammed, S.-E.A. (1984). *Stochastic functional differential equations.* Pitman, Boston.

Mohammed, S.-E. A. and Scheutzow, M. and von Weizsaecker, H. (1986). Hyperbolic state space decomposition for a linear stochastic delay equation. *SIAM J. Control Optimization*, 24:543–551.

Mohammed, Salah-Eldin A. and Scheutzow, Michael K.R. (1990). Lyapunov exponents and stationary solutions for affine stochastic delay equations. *Stochastics Stochastics Rep.*, 29(2):259–283.

Neumann, M. and von Sachs, R. (1995). Wavelet thresholding: Beyond the Gaussian i.i.d. situation. In Antoniadis, A., editor, *Wavelets and statistics. Proceedings of the 15th French-Belgian meeting of statisticians, 1994.* Springer, New York.

Nussbaum, Michael and Pereverzev, Sergej V. (1999). The degrees of ill-posedness in stochastic and deterministic noise models. Preprint 509, Weierstraß-Institut Berlin (WIAS).

Pinsker, M.S. (1980). Optimal filtration of square-integrable signals in Gaussian noise. *Probl. Inf. Transm.*, 16:120–133.

Protter, Philip (1992). *Stochastic integration and differential equations. A new approach. 2nd corr. printing.* Springer, Berlin.

Putschke, Maren Ulrike (2001). *Stochastische Funktionaldifferentialgleichungen und lokal-asymptotische Eigenschaften ihrer Parameterschätzungen.* PhD thesis, Humboldt Universität zu Berlin.

Revuz, Daniel and Yor, Marc (1999). *Continuous martingales and Brownian motion. 3rd ed.* Springer, Berlin.

Riedle, Markus (2001). Affine stochastic differential equations with infinite delay on abstract path spaces. Discussion paper, Humboldt-Universität Berlin, Sonderforschungsbereich 373. to be published.

Rothkirch, Ulf (1993). *Schätzung von Parametern in linearen stochastischen Funktionaldifferentialgleichungen.* PhD thesis, Universität Jena.

Rudin, Walter (1991). *Functional analysis. 2nd ed.* McGraw-Hill, New York.

Sakhnovich, L.A. (1996). *Integral equations with difference kernels on finite intervals.* Birkhäuser, Basel.

Scheutzow, Michael (1983). *Qualitatives Verhalten der Lösungen von eindimensionalen nichtlinearen stochastischen Differentialgleichungen mit Gedächtnis.* PhD thesis, Universität Kaiserslautern.

Scheutzow, M. (1984). Qualitative behaviour of stochastic delay equations with a bounded memory. *Stochastics*, 12:41–80.

Schumaker, Larry L. (1981). *Spline functions: basic theory.* Wiley-Interscience, New York.

Spokoiny, V.G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Stat.*, 24(6):2477–2498.

Triebel, Hans (1983). *Theory of function spaces.* Mathematik und ihre Anwendungen in Physik und Technik, Bd. 38. Akademische Verlagsgesellschaft, Leipzig.

Vakhaniya, N.N. and Tarieladze, V.I. and Chobanyan, S.A. (1987). *Probability distributions on Banach spaces.* Reidel, Dordrecht.

Verduyn Lunel, S. (2000). Parameter identifiability of differential delay equations. Preprint 14, Mathematisch Instituut, Universiteit Leiden.

Zeidler, Eberhard (1995). *Applied functional analysis. Main principles and their applications.* Applied Mathematical Sciences. 109. Springer-Verlag, New York.

Lebenslauf

Markus Reiß

- geboren am 22.5.1973 in Berlin, Deutschland
- unverheiratet
- Grundschule: 1979-1983, Berlin
- Gymnasium: 1983-1992, Berlin
- Zivildienst: 1993-1994, Berlin
- Ablauf des Mathematikstudiums:
  - 1992-1993: Freie Universität, Berlin
  - 1994-1995: Freie Universität, Berlin
  - 1995-1996: Somerville College, Oxford
  - 1996-1999: Freie Universität, Berlin
  - 1999: Diplom mit einer Arbeit zum Thema *Wavelets und Sobolevräume*
- Stelle als wissenschaftlicher Mitarbeiter am Institut für Mathematik der Humboldt-Universität zu Berlin, seit 1999
- Forschungsaufenthalt am Laboratoire de probabilités et modèles aléatoires, Université de Paris 6, Oktober 2000 bis März 2001